

ИНФОРМАЦИОННОЕ ОБЩЕСТВО

4-5
2021

Большие данные для цифровой экономики

Мониторинг развития и использования технологий
работы с большими данными

Индустрия работы с большими данными

Использование больших данных в различных сферах
деятельности

Воздействие больших данных на экономику

Факторы, влияющие на работу с большими данными

Стандарты для работы с большими данными

№ 4-5 2021

ИНФОРМАЦИОННОЕ ОБЩЕСТВО

ОСНОВАН В 1989 ГОДУ
ВЫХОДИТ 6 РАЗ В ГОД

УЧРЕДИТЕЛИ:

ИНСТИТУТ РАЗВИТИЯ ИНФОРМАЦИОННОГО ОБЩЕСТВА
РОССИЙСКАЯ ИНЖЕНЕРНАЯ АКАДЕМИЯ

ГЛАВНЫЙ РЕДАКТОР: РЕДАКЦИОННЫЙ СОВЕТ:

ЕРШОВА Татьяна
Викторовна – канд.
экон. наук

ХОХЛОВ Юрий Евгеньевич (председатель) – канд. физ.–мат. наук, доц., акад. РИА
ОРЛОВ Степан Владимирович (зам. председателя) – канд. экон. наук
АЛЕКСЕЕВА Ирина Юрьевна – д-р филос. наук, доц.
БОГДАНОВ Александр Владимирович – д-р физ.–мат. наук, проф.
ВАРТАНОВА Елена Леонидовна – д-р фил. наук, проф.
ВЕРШИНСКАЯ Ольга Николаевна – д-р экон. наук
ВОЙСКУНСКИЙ Александр Евгеньевич – д-р психол. наук
ДЕЖИНА Ирина Геннадьевна – д-р экон. наук, проф.
ЕЛИЗАРОВ Александр Михайлович – д-р физ.–мат. наук, проф.
ИВАНОВ Алексей Дмитриевич – д-р экон. наук, чл.-кор. РАЕН
ИВАХНЕНКО Евгений Николаевич – д-р филос. наук, проф.
КОГАНОВСКИЙ Михаил Рувимович – канд. техн. наук, доц.
КОЛИН Константин Константинович – д-р техн. наук, проф., засл. деятель науки РФ
КУЗНЕЦОВА Наталия Ивановна – д-р филос. наук, проф.
МЕНДЖКОВИЧ Андрей Семенович – д-р химических наук, ст. науч. сотрудник
ОЛЕЙНИК Андрей Владимирович – д-р техн. наук, проф.
РАЙКОВ Александр Николаевич – д-р техн. наук, проф.
РУСАКОВ Александр Ильич – д-р хим. наук, проф.
СЕМЕНОВ Алексей Львович – д-р физ.–мат. наук, акад. РАН, действ. член РАО
СЕМЕНОВ Евгений Васильевич – д-р филос. наук, проф.
СЕРДЮК Владимир Александрович – канд. техн. наук, доц.
СМОЛЯН Георгий Львович – д-р филос. наук, проф.
СТРЕЛЬЦОВ Анатолий Александрович – д-р техн. наук, д-р юрид. наук, проф., засл. деятель науки РФ
ТАТАРОВА Галина Галеевна – д-р социол. наук, проф.
ЧЕРЕШКИН Дмитрий Семенович – д-р техн. наук, проф., акад. РАЕН
ШАПОШНИК Сергей Борисович
ЩУР Лев Николаевич – д-р физ.–мат. наук, проф.
ЯКУШЕВ Михаил Владимирович

Журнал зарегистрирован в Роспечати
(Per № 015 766 от 01.07.1999)
ISSN 1605-9921 (эл.)

Адрес редакции: Москва, Армянский переулок,
д. 9, офис 310
Для переписки: 101000, Москва, Главпочтамт, а/я 716
Тел.: +7 (495) 912-22-29
Электронная почта: info@infosoc.iis.ru
Веб-сайт: www.infosoc.iis.ru

Позиция редакции может не совпадать с мнением авторов.

Авторы несут ответственность за патентную чистоту, достоверность и точность приведенных фактов, цитат, экономико-статистических данных, собственных имен, географических названий и прочих сведений, а также за разглашение данных, не подлежащих открытой публикации. При любом использовании оригинальных материалов ссылка на журнал обязательна.

ЛЕГАЛЬНЫЙ ПОЛЬЗОВАТЕЛЬ
Пара(-)Тайп
IN LEGAL USE

В макете журнала использованы шрифты
ООО НПЦ «ПараТайп»

**ПУБЛИКУЕМЫЕ МАТЕРИАЛЫ ПРОШЛИ ПРОЦЕДУРУ
РЕЦЕНЗИРОВАНИЯ И ЭКСПЕРТНОГО ОТБОРА**

**ЖУРНАЛ ВКЛЮЧЕН В ПЕРЕЧЕНЬ ПЕРИОДИЧЕСКИХ ИЗДАНИЙ, РЕКОМЕНДОВАННЫХ ВЫСШЕЙ АТТЕСТАЦИОННОЙ КОМИССИЕЙ
РФ ДЛЯ ПУБЛИКАЦИИ МАТЕРИАЛОВ КАНДИДАТСКИХ И ДОКТОРСКИХ ДИССЕРТАЦИОННЫХ ИССЛЕДОВАНИЙ. ЖУРНАЛ ВХОДИТ В
ДАННЫЙ СПИСОК С 26 ФЕВРАЛЯ 2010 ГОДА.
С 2015 ГОДА ЖУРНАЛ ВХОДИТ В РОССИЙСКУЮ ПОЛКУ ЖУРНАЛОВ (RUSSIAN SCIENCE CITATION INDEX) НА ПЛАТФОРМЕ WEB OF
SCIENCE.**

© Институт развития информационного общества, 2021

Публикации в журнале «Информационное общество» доступны в открытом доступе по международной лицензии
Creative Commons «С указанием авторства - Некоммерческая - С сохранением условий» версии 4.0 Международная

СОДЕРЖАНИЕ № 4-5 2021

СЛОВО РЕДАКТОРА

- 1 ХОХЛОВ Юрий Евгеньевич **Большие данные для цифровой экономики**

ИЗМЕРЕНИЕ ИНФОРМАЦИОННОГО ОБЩЕСТВА

- 2 ЕРШОВА Татьяна Викторовна, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович
Методология мониторинга развития и использования технологий работы с большими данными

ИНФОРМАЦИОННОЕ ОБЩЕСТВО: ПОЛИТИКА И ФАКТОРЫ РАЗВИТИЯ

- 33 ОРЛОВ Степан Владимирович, ПАДЖЕВ Валентин Валентинович, ХОХЛОВ Юрий Евгеньевич
Государственная политика и регулирование работы с большими данными
- 53 ЕЛИЗАРОВ Александр Михайлович, ПАДЖЕВ Валентин Валентинович, ХОХЛОВ Юрий Евгеньевич
Система управления и механизмы финансирования работы с большими данными
- 66 ШАПОШНИК Сергей Борисович, ЯНЫШЕН Анна Андреевна
Человеческий капитал для работы с большими данными в Российской Федерации
- 90 ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович
Исследования и разработки в области работы с большими данными
- 110 ЕРШОВ Петр Сергеевич, ХОХЛОВ Юрий Евгеньевич
Цифровая инфраструктура для работы с большими данными

СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ АСПЕКТЫ ИНФОРМАЦИОННОГО ОБЩЕСТВА

- 132 МАЛАХОВ Вадим Александрович, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович, ЮРЕВИЧ Максим Андреевич
Большие данные: социальные и экономические эффекты

ИНФОРМАЦИОННОЕ ОБЩЕСТВО И ВЛАСТЬ

- 150 КАТИН Александр Владимирович, ХОХЛОВ Юрий Евгеньевич
Мониторинг использования технологий работы с большими данными в системе государственного управления в России

ОБРАЗОВАНИЕ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 166 ЕЛИЗАРОВ Александр Михайлович, ХОХЛОВ Юрий Евгеньевич
Мониторинг использования технологий работы с большими данными в российской образовательной системе

ЗДРАВООХРАНЕНИЕ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 185 КАТИН Александр Владимирович, ХОХЛОВ Юрий Евгеньевич
Мониторинг использования технологий работы с большими данными в российской системе здравоохранения

НАУКА И ИННОВАЦИИ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 200 МАЛАХОВ Вадим Александрович, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович **Использование технологий работы с большими данными в российской науке**

ИНФОРМАЦИОННОЕ ОБЩЕСТВО И ПРАВО

- 220 АВЕРКИН Алексей Николаевич, АФНАСЬЕВ Сергей Дмитриевич, МИКРЮКОВ Андрей Александрович, ПАДЖЕВ Валентин Валентинович, РАЙКОВ Александр Николаевич, ХОХЛОВ Юрий Евгеньевич, ХРАМЦОВСКАЯ Наталья Александровна **Стандартизация работы с большими данными: международные и национальные стандарты**

ТЕХНОЛОГИИ ИНФОРМАЦИОННОГО ОБЩЕСТВА

- 259 ЕРШОВ Петр Сергеевич, КАТИН Александр Владимирович, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович **Модель BD4DE-MM зрелости работы с большими данными в организации**
- 278 МАЛАХОВ Вадим Александрович, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович **Индустрия работы с большими данными**
- 300 ЕРШОВ Петр Сергеевич, ХОХЛОВ Юрий Евгеньевич, ШАПОШНИК Сергей Борисович **Использование технологий работы с большими данными в российском бизнесе**

ДОВЕРИЕ И БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 315 КАТИН Александр Владимирович, ХОХЛОВ Юрий Евгеньевич **Доверие и безопасность в работе с большими данными в России**

ЦИФРОВОЕ СЕЛЬСКОЕ ХОЗЯЙСТВО

- 334 ОТМАХОВА Юлия Сергеевна, ДЕВЯТКИН Дмитрий Алексеевич, УСЕНКО Наталья Ивановна **Анализ цифровых технологий в агропродовольственной сфере с использованием методов обработки больших данных**

Слово редактора

БОЛЬШИЕ ДАННЫЕ ДЛЯ ЦИФРОВОЙ ЭКОНОМИКИ**Хохлов Юрий Евгеньевич***Кандидат физико-математических наук, доцент**Институт развития информационного общества, председатель Совета директоров**РЭУ имени Г.В. Плеханова, научный руководитель Базовой кафедры цифровой экономики ИРИО**Москва, Российская Федерация**yuri.hohlov@iis.ru*

Данные повсюду. И у них есть ценность. Осознание этого пришло к человечеству не так давно, на волне цифровой революции. Переход «в цифру» привел к тому, что нас захлестнул девятый вал данных в цифровой форме, в которых спрятана информация, необходимая для принятия осознанных, продуманных решений. Этот переход не состоялся бы без бурного развития телекоммуникационной инфраструктуры, «убившей» расстояние между людьми и организациями, которые хотели общаться, и вычислительной инфраструктуры, «убившей» время, которое человек ранее тратил на свои жизненно важные расчеты.

Феномен «больших данных» насчитывает от силы два десятка лет, но за это время подрывные технологии работы с большими данными успели пройти путь от зарождения до хайпа и переключиться на плато продуктивности, где их применение стало обычным, рутинным делом, приносящим умелым пользователям явные экономические и социальные дивиденды.

Настоящий выпуск журнала «Информационное общество» содержит результаты трехлетнего проекта Национальной технологической инициативы, направленного на создание системы мониторинга (и стандартизации) процессов производства, использования многочисленных технологий работы с большими данными и их воздействия на социально-экономическое развитие. Этот проект был реализован в рамках программы Центра НТИ по большим данным при Московском университете и явил собой яркий пример многостороннего партнерства организаций, представляющих основные заинтересованные стороны в цифровом развитии: органы власти, бизнес, гражданское общество, научно-образовательное сообщество.

Среди публикаций данного выпуска наш читатель найдет как теоретические исследования по построению концептуальной схемы мониторинга BD4DE (Big Data for Digital Economy) или модели зрелости BD4DE-MM работы с большими данными в организации, так и практические результаты оценки состояния дел с использованием больших данных в конкретных сферах деятельности: бизнесе или государственном управлении, здравоохранении или образовании. Он сможет увидеть, какие факторы влияют на то, чтобы работа с большими данными приносила плоды; он также сможет понять, от чего это зависит – от государственной ли политики, от уровня ли подготовки кадров, от наличия ли в стране исследовательского потенциала или разработчиков отечественных технологий. При этом ему станет ясно, что невозможно целенаправленно развивать сложную социально-технологическую систему, не имея петли обратной связи в виде системы мониторинга, которая позволяет осуществлять управляющие воздействия.

Конечно, детальные ответы на эти и другие вопросы можно получить лишь на основе полноценных масштабных опросов организаций и экспертов, которые были невозможны в рамках бюджета исследовательского проекта, но проведенное авторами пилотирование разработанной системы мониторинга вселяет уверенность, что разработанный инструментарий найдет свое применение для развития цифровой экономики страны в целом, для цифровой трансформации различных сфер деятельности, а также отдельных организаций. И результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» служат решению этой задачи.

© Хохлов Ю.Е., 2021

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_01

Измерение информационного общества

МЕТОДОЛОГИЯ МОНИТОРИНГА РАЗВИТИЯ И ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Ершова Татьяна Викторовна

*Кандидат экономических наук
Институт развития информационного общества, генеральный директор
Научно-аналитический журнал «Информационное общество», главный редактор
Москва, Российская Федерация
tatiana.ershova@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Аннотация

Предложены концептуальная схема BD4DE (Big Data for Digital Economy) и методология мониторинга использования технологий работы с большими данными в цифровой экономике Российской Федерации. Проведен анализ многочисленных концептуальных схем мониторинга развития и использования цифровых технологий для социально-экономического развития (от оценок «электронной готовности» и развития информационного общества до оценок уровня цифровой зрелости и измерения цифровой экономики). Концептуальная схема BD4DE охватывает предметные области мониторинга: производство, использование и воздействие технологий работы с большими данными на социально-экономическое развитие, а также факторы, которые на это влияют (государственная политика и регулирование, институты управления и финансирования, цифровая инфраструктура, человеческий капитал, исследования и разработки, информационная безопасность). Даны рекомендации по формированию системы измеримых показателей, характеризующих предложенные предметные области, обозначены источники, методы и инструменты для сбора необходимых данных.

Ключевые слова

большие данные; технологии работы с большими данными; мониторинг развития и использования цифровых технологий; концептуальная схема мониторинга; Big Data for Digital Economy; BD4DE; методология мониторинга и оценки; производство технологий работы с большими данными; использование больших данных; воздействие технологий работы с большими данными; факторы влияющие на работу с большими данными

© Ершова Т.В., Хохлов Ю.Е., Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_02

Введение

Стремительные социально-экономические изменения, происходящие в развитых и большинстве развивающихся стран, сегодня связывают с масштабным использованием цифровых технологий «третьей волны» таких как технологии искусственного интеллекта, аналитики больших данных, распределенного реестра, робототехники, аддитивного производства, виртуальной/дополненной реальности, интернета вещей [1]. Процессы цифровой трансформации экономики и общества [2] порождают как большие ожидания (экономического роста, улучшения качества услуг и др.), так и опасения (сокращения рабочих мест, усиления неравенства, роста угроз информационной безопасности) [3]. В ответ на происходящие изменения во многих странах, а также в России, приняты или разрабатываются цифровые стратегии и планы действий, нацеленные на реализацию возможностей новой волны цифровых технологий и парирование рисков, связанных с их использованием. Все это делает актуальной задачу формирования системы управления этими процессами и ее научно-информационного обеспечения.

Одним из важнейших элементов управления любой социальной системой является механизм обратной связи, который применительно к цифровой трансформации реализуется в виде комплексной системы мониторинга развития, использования и воздействия цифровых технологий на национальном, региональном и отраслевом уровнях [4].

Вместе с тем, несмотря на всеобщее признание, что данные в цифровой форме, их анализ и использование являются сегодня одной из основ цифровой экономики [5], а технологии хранения и анализа больших данных – одним из драйверов современного этапа цифровой трансформации различных сфер деятельности, в Российской Федерации сохраняется дефицит сведений об этой сфере технологического развития. Этот дефицит лишь отчасти был преодолен в 2021 г., когда в России были опубликованы данные, полученные на основе модернизированной в 2020 г. основной формы федерального статистического наблюдения за использованием цифровых технологий (№3-информ), в которую было включено несколько вопросов об использовании технологий работы с большими данными [6]. Обновленная форма частично гармонизирована с последними рекомендациями и модельными анкетами международных организаций (ОЭСР и Евростата), в которые еще в 2014–2020 гг. были включены показатели использования аналитики больших данных в бизнесе [7, 8, 9, 10]. Но проведенная модернизация статистического наблюдения – малая часть решения проблемы. Доступные данные не позволяют проводить комплексный мониторинг развития и использования технологий работы с большими данными в экономике страны.

Развитие и широкомасштабное использование технологий хранения и анализа больших данных – относительно новая, быстро развивающаяся (в плане технологий, основанных на них решений и способов их использования) и достаточно специфичная область. С этим связан ряд объективных трудностей получения необходимой мониторинговой информации:

- технологии хранения и анализа больших данных начинают активно использоваться не только в бизнесе, они являются сегодня технологическим драйвером изменений и в других сферах деятельности (государственном и муниципальном управлении, здравоохранении, образовании, науке), по которым отсутствуют международные стандарты и национальная система мониторинга использования технологий работы с большими данными;
- действующие международные и гармонизированные с ними отечественные классификаторы и группировки (видов экономической деятельности, товаров и услуг, кодов патентной классификации, специальностей и направлений подготовки, областей научных исследований, цифровых технологий), не отражают даже в минимально необходимой детализации, исследования и разработки, подготовку кадров, производство и продажу товаров и услуг, а также другие виды деятельности, связанные с технологиями хранения и анализа больших данных, что не позволяет осуществлять мониторинг с опорой на действующие системы сбора данных.

В этой ситуации разработка и реализация предметно ориентированной системы мониторинга, включающей концептуальную схему, набор показателей, классификаторы, инструменты измерения и собственную систему регулярных обследований и сбора данных, является актуальной, но при этом достаточно сложной и наукоемкой задачей.

В данной статье представлены основные результаты разработки методологии мониторинга развития и использования технологий работы с большими данными (Big Data for Digital Economy, BD4DE), включая концептуальную схему мониторинга, краткое описание разработанных и

используемых классификаторов, методов и инструментария сбора исходных данных. Более детальное описание разработанной методологии представлено в других статьях данного номера журнала, посвященных отдельным предметным областям мониторинга.

1 Обзор литературы

Международные стандарты и системы мониторинга развития информационного общества и цифровой экономики, а также действующие классификаторы пока не адаптированы для целей измерения такой достаточно специфической предметной области как развитие и использование технологий работы с большими данными. В этом направлении сделаны только первые шаги - в последние годы международные организации и объединения (в частности ОЭСР и ЕС) для учета тенденций технологического развития обновили стандарты статистического наблюдения за производством и использованием цифровых технологий, включив туда несколько показателей использования аналитики больших данных.

Поэтому, выстраивая комплексную систему мониторинга этой предметной области, необходимо учитывать и использовать накопленный опыт и подходы, разработанные для измерения более широкой области - развития и использования цифровых технологий.

Разработка системы мониторинга начинается с определения концептуальных схем (рамоч) мониторинга и оценки. В данном разделе представлен краткий аналитический обзор концептуальных схем различных систем мониторинга и оценки использования цифровых технологий, включая методологии оценки «электронной готовности» (e-readiness assessment), композитные индексы развития цифровой экономики и информационного общества, а также международные стандарты статистического наблюдения в этой области. Кроме того, описаны первые подходы на международном уровне к мониторингу и оценке использования и воздействия технологий хранения и анализа больших данных.

1.1 Концептуальные схемы оценки электронной готовности

Разработка методологий оценки электронной готовности и их применение начались в конце 1990-х годов, что во многом было связано с двумя обстоятельствами: (1) в развитых странах стали очевидны масштабы и потенциал использования ИКТ (таких как персональные компьютеры, интернет, а также основанные на них сервисы и технологии) для развития различных сфер деятельности; (2) образовался существенный разрыв между развивающимися и развитыми странами в уровне «электронного развития», который получил название информационного (или цифрового) неравенства. В этой ситуации использование ИКТ для социально-экономического развития стало важной частью повестки дня для правительств многих стран и международных организаций и стало дополнительным стимулом для разработки информационно-аналитических инструментов обеспечения политики «электронного развития» и преодоления цифрового разрыва.

В общей сложности было разработано больше десяти методологий оценки уровня электронной готовности, которые отражали взгляды и интересы, методы и подходы национальных и международных исследовательских центров, консалтинговых фирм, правительственных организаций.

Методологии оценки электронной готовности различались, но можно выделить общее понимание готовности как наличие условий для использования ИКТ для развития страны, региона или отрасли экономики. При этом оценка уровня готовности предполагала наличие метрик, критериев и шкал для оценки степени в которой страна, регион или отрасль экономики готовы получать выгоды, возникающие в результате использования ИКТ [11].

Разработанные подходы к оценке электронной готовности отличались наборами предметных областей и характеризующих их показателей, методами и шкалами квантификации оценки, инструментарием для измерения показателей. Наличие достаточно большого числа обзоров [11-14, 21] избавляет от необходимости разбирать детали и позволяет сосредоточиться на релевантных для данной работы аспектах этих методологий, а именно концептуальных схемах оценки электронной готовности.

Согласно приведенному выше определению, основным содержанием (фокусом) оценки электронной готовности была оценка наличия условий для широкомасштабного распространения и использования ИКТ. Ряд методологий ограничивались именно этим аспектом - анализом факторов, влияющих на процессы электронного развития, в других присутствовали показатели,

характеризующие использование ИКТ в различных сферах деятельности (электронное правительство, электронная коммерция, электронное образование и др.). Приведем несколько примеров.

1. Первый инструмент для оценки готовности к электронным операциям был разработан в 1998 г. (опубликован в 2000 г.) в рамках проекта «Политика в области компьютерных систем» (Computer Systems Policy Project – CSPP) и назывался «Руководство по готовности для жизни в сетевом мире» [15]. Руководство позволяло измерить распространенность и использование ИКТ в домах, школах, на предприятиях, в медицинских и государственных учреждениях, с дополнительным акцентом на конкуренцию среди поставщиков услуг доступа к интернету, скорость доступа и государственную политику. Измерения проводились в пяти предметных областях:

- инфраструктура;
- доступ;
- приложения и услуги;
- экономика;
- «средства обеспечения» (политика, конфиденциальность, безопасность, повсеместность).

2. Методология Центра международного развития Гарвардского университета «Готовность к сетевому миру: руководство для развивающихся стран» (2000 г.) [16]. Данное руководство предоставило инструменты для измерения 19 различных категорий, включая доступность; скорость и качество доступа к сети; использование ИКТ в школах, на рабочем месте, в экономике, в системе государственного управления и в повседневной жизни; политику в области ИКТ (телекоммуникации и торговля); программы обучения в области ИКТ; разнообразие поставщиков контента, представленных в интернете. Методология Гарвардского университета была одной из первых публично открытых методологий.

3. Еще одна публичная, готовая к использованию (в т.ч. для самооценки) методология – оценка готовности к электронной торговле организации Азиатско-Тихоокеанского экономического сотрудничества (АТЭС, 2000 г.) [17], которая предполагала проведение измерений в шести предметных областях:

- базовая инфраструктура и технологии (скорость, цены, доступ, рыночная конкуренция, отраслевые стандарты и иностранные инвестиции);
- доступ к сетевым услугам (пропускная способность, отраслевое разнообразие, экспортный контроль, регулирование кредитных карт);
- использование интернета (в бизнесе, государственном секторе, дома);
- продвижение и содействие (отраслевые стандарты);
- навыки и человеческие ресурсы (образование в сфере ИКТ, рабочая сила);
- позиционирование страны для развития цифровой экономики (налоги и тарифы, саморегулирование отрасли, государственное регулирование, доверие потребителей).

4. Характерным примером методологии оценки готовности различных стран, производимой сторонними организациями по своим разработкам, является методология, представленная в докладе по готовности к электронному бизнесу компании McConnell International [18]. В докладе оценка делалась по пяти областям:

- связь (инфраструктура, доступ и цены);
- электронное лидерство (государственная политика и регулирование);
- информационная безопасность (интеллектуальная собственность, конфиденциальность, электронные подписи);
- человеческий капитал (образование в области ИКТ, наличие квалифицированной рабочей силы);
- бизнес-климат для электронного бизнеса (конкуренция, политическая и финансовая стабильность, иностранные инвестиции, финансовая инфраструктура).

5. Первым отечественным исследованием в данном направлении стала оценка готовности России к развитию информационного общества [19], выполненная Институтом развития информационного общества в 2001 г. в ходе реализации проекта «Российский портал развития». Следующим шагом стала подготовка аналитического доклада «Готовность России к информационному обществу» [20], который был поддержан специальным грантом программы «infoDev» Всемирного банка. Для оценки уровня готовности Российской Федерации к

широкомасштабному использованию ИКТ для социально-экономического развития была переработана и адаптирована методология оценки готовности к сетевому миру Центра международного развития Гарвардского университета [16].

Оценка электронной готовности России проводилась по 8 компонентам: 4 группам факторов, влияющих на электронное развитие (доступ к ИКТ, человеческий капитал, бизнес-климат, государственное регулирование) и 4 предметным областям использования ИКТ (электронное правительство, ИКТ в образовании, ИКТ в культуре, электронный бизнес), см. рисунок 1.



Рис. 1. Концептуальные рамки оценки готовности России к информационному обществу
Источник: [20]

К 2003 г., когда был опубликован один из первых аналитических обзоров методологий оценки электронной готовности, уже около 137 стран были оценены по меньшей мере по одной методике, 55 стран были оценены как минимум пять раз различными организациями, а 10 стран были оценены более 8 раз [21].

Подводя итоги развития этого направления исследований, можно сказать, что в большинстве разработанных методологий оценки электронной готовности присутствуют следующие основные предметные области оценки и соответствующие им показатели:

- ИКТ-инфраструктура (техническая доступность – покрытие сотовой связью, стационарной связью, доступ к интернету; ценовая доступность; качество связи; пропускная способность каналов связи и т.д.);
- человеческий капитал (навыки использования ИКТ, ИКТ-специалисты, подготовка кадров);
- государственная политика и регулирование (программы, законодательство и стандарты в сфере ИКТ);
- деловой или инвестиционный климат (доступность финансовых ресурсов; административные барьеры; транспортная, логистическая и иная инфраструктура для ведения бизнеса; уровень конкуренции и др.);
- информационная безопасность и доверие (технические средства защиты, инциденты, доверие потребителей и т.д.);
- использование ИКТ в ключевых сферах деятельности (государственное управление, образование, здравоохранение, бизнес, домохозяйства и население, другие сферы деятельности).

Отметим, что серьезным недостатком концептуальных схем первого поколения методологий оценки электронной готовности было практически полное отсутствие показателей, характеризующих воздействие ИКТ, то есть социально-экономические эффекты от их использования.

1.2 Индексы развития информационного общества и цифровой экономики

Отметим, что отдельное рассмотрение композитных индексов развития информационного общества (и цифровой экономики) носит достаточно условный характер. Некоторые из индексов были разработаны на основе методологий оценки электронного развития, определяющих на основе порядковых (четырёх-, пятибалльных и более) шкал уровень готовности в отдельных областях по определенным критериям. Другие индексы сразу конструировались в рамках идеологии оценки электронной готовности, поэтому в большинстве первых индексов в названии присутствовало слово «готовность», которое в настоящее время практически везде исключено.

Для индексов, оценивающих конкретные направления развития и использования ИКТ (электронное правительство, электронный бизнес, развитие ИКТ), также характерно включение в концептуальную схему оценки факторов, влияющих на их использование. Так, в концептуальной схеме Индекса развития ИКТ [22, 23] Международного союза электросвязи (МСЭ) присутствует оценка человеческого капитала, а в Индекс развития электронного правительства [24] Департамента по экономическим и социальным вопросам ООН включен индекс телекоммуникаций, характеризующий состояние ИКТ-инфраструктуры (проникновение мобильной связи, фиксированного и мобильного ШПД и другие) (см. рисунок 2).

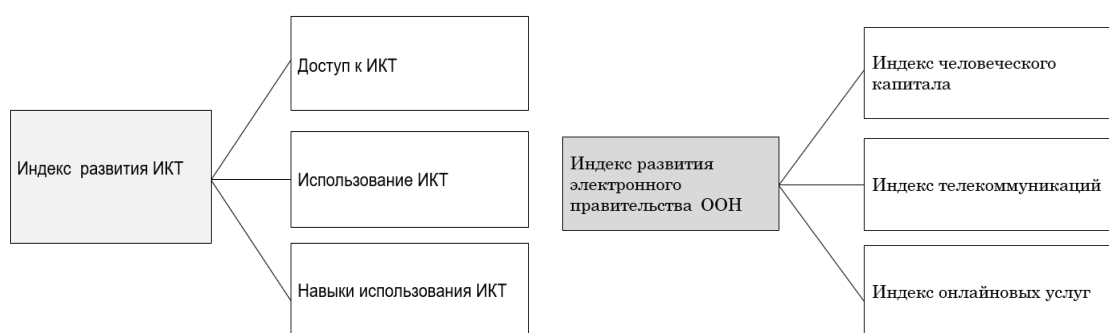


Рис. 2. Концептуальная схема Индекса развития ИКТ (МСЭ) и Индекса развития электронного правительства (ДЭСВ ООН)
Источники [22, 23], [24]

В Европейском союзе в последние годы используется комплексный Индекс цифровой экономики и общества (Digital Economy and Society Index, DESI) [25], в котором выделены шесть направлений (размерностей) оценки:

- подключенность (connectivity) – характеризует уровень ИКТ-инфраструктуры;
- человеческий капитал (human capital) – оценивает цифровые навыки населения;
- использование онлайн-сервисов (use of internet services) – оценивает использование населением онлайн-сервисов для различных видов деятельности;
- интеграция цифровых технологий (integration of digital technology) – оценивает использование цифровых технологий бизнесом;
- цифровые государственные услуги (digital public services) – включает показатели использования онлайн-государственных услуг;
- ИКТ-сектор и его активность в сфере исследований и разработок (ICT Sector and its R&D Performance) – оценивает уровень развития и инновационную активность ИКТ-сектора.

В расширенном индексе цифровой экономики и общества International-DESI проводятся международные сопоставления стран-участниц с 17-ю странами, не входящими в ЕС [26].

Наиболее детально факторы использования ИКТ были представлены в Индексе готовности к сетевому миру Всемирного экономического форума, который изначально разрабатывался на основе методологии оценки готовности Центра международного развития Гарвардского университета. Индекс готовился сначала в сотрудничестве с этим центром, а затем со Всемирным банком, бизнес-школой Insead и Корнеллским университетом и имел следующую структуру [27].

1. Подындекс среды.

- Политическая и нормативная среда.
- Бизнес-среда и инновационная среда.

2. Подындекс готовности.

- Инфраструктура и цифровой контент.
- Доступность.
- Навыки.

3. Подындекс использования.

- Использование населением.
- Использование бизнесом.
- Использование государством.

4. Подындекс воздействия.

- Экономические эффекты.
- Социальные эффекты.

Начиная с 2019 г. обновленный Индекс готовности к сетевому миру рассчитывается независимым научно-образовательным Институтом Портуланса [28] и имеет следующую структуру (см. рисунок 3).

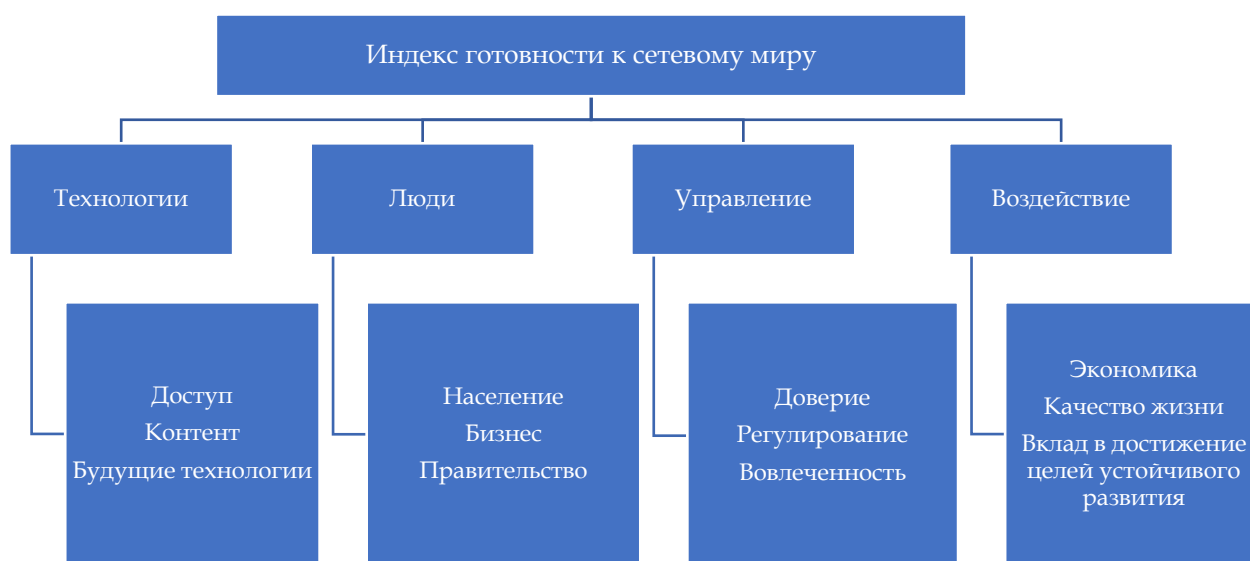


Рис. 3. Концептуальная схема Индекса готовности к сетевому миру 2020
Источник [28]

В Российской Федерации комплексный Индекс готовности регионов России к информационному обществу был разработан и рассчитывался начиная с 2005 г. сотрудниками Института развития информационного общества [29-35]. Индекс российских регионов строится на показателях, характеризующих три ключевых фактора электронного развития (человеческий капитал, экономическая среда, ИКТ-инфраструктура) и показателей доступа и использования ИКТ в шести сферах деятельности (государственном и муниципальном управлении, бизнесе, образовании, здравоохранении, культуре), а также на показателях использования ИКТ домохозяйствами и населением. Индекс также позволяет измерять различия между субъектами Российской Федерации по интегральным показателям информационного развития [36].

В 2018 г. экспертами Института развития информационного общества, Национального центра цифровой экономики МГУ имени М.В.Ломоносова и АО «Гринатом» была разработана методология расчета Национального индекса развития цифровой экономики и проведены пилотные расчеты, опубликованные в докладе «Национальный индекс развития цифровой экономики: Пилотная реализация» [37]. В основу концептуальной схемы этого индекса (см. рисунок 4) положен подход, согласно которому потенциал цифровых технологий для социально-экономического развития может быть реализован только при наличии стимулирующего развития нормативного регулирования, развитого человеческого капитала для цифровой экономики,

благоприятного делового климата, эффективной научно-инновационной системы, безопасной цифровой инфраструктуры и конкурентоспособного цифрового сектора экономики. При формировании предметных областей для измерения развития цифровой экономики использовался системный подход, предполагающий оценку условий (факторов) развития цифровой экономики, процессов цифровой трансформации различных отраслей экономики, а также социально-экономических эффектов использования цифровых технологий.



Рис. 4. Концептуальная схема Национального индекса развития цифровой экономики
Источник: [37]

В 2016–2017 гг. Минкомсвязи России представило два выпуска рейтинга субъектов Российской Федерации по уровню развития информационного общества (разрабатывался подведомственной организацией министерства – Центром экспертизы и координации информатизации). Для построения рейтинга в 2016 г. использовалось 58 показателей по семи подындаксам [38]. На состоявшемся 13 октября 2017 г. совете по региональной информатизации Минкомсвязь представила рейтинг регионов по развитию информационного общества в РФ на 2017 г. [39], методика которого уже включала оценку по 120 показателям и 17 подындаксам. Рейтинг был разработан в рамках реализации Концепции региональной информатизации и основан на показателях, характеризующих основные направления реализации Концепции и состояние факторов развития информационного общества в регионах. В состав композитного индекса, на основе которого строился рейтинг, в 2017 г. входило 2 интегральных индекс-компонента, объединяющих 17 подындаксов:

- факторы развития информационного общества (подындаксы: человеческий капитал, ИКТ-инфраструктура, управление информатизацией);
- использование ИКТ для развития (подындаксы: электронное правительство, использование ИКТ в домохозяйствах и населением, ИКТ в сфере культуры, ИКТ в сфере предпринимательства и торговли, ИКТ в сфере здравоохранения, ИКТ в сфере образования, ИКТ в сфере транспорта, ИКТ в сфере строительства, ИКТ в сфере обеспечения безопасности жизнедеятельности, ИКТ в сфере сельского хозяйства, ИКТ в сфере государственных и муниципальных финансов, ИКТ в сфере социальной защиты населения, ИКТ в сфере труда и занятости, ИКТ в сфере энергетики).

Сравнительный анализ перечисленных индексов развития информационного общества и цифровой экономики приводит к следующим выводам.

1. Концептуальные схемы большинства индексов сформированы в рамках подхода, заложенного в оценках электронной готовности – наборы показателей, из которых строится интегральный индикатор, группируются по направлениям, связанным с формированием условий для развития и использования ИКТ: человеческий капитал, ИКТ-инфраструктура, государственная политика и регулирование, деловая среда и др., а также, в ряде случаев, по областям использования (электронное правительство, электронный бизнес и др.).

2. В существенно больших масштабах, чем в первом поколении методологий оценки электронной готовности, в современных комплексных индексах используются показатели, характеризующие ИКТ-сектор, НИОКР и инновации, а также социально-экономических эффекты от использования цифровых технологий. Последнее направление представляется особенно значимым, так как ориентировано на оценку воздействия технологий, которые связаны с целевыми установками и рисками цифрового развития. Наиболее развернуто это направление представлено в концептуальных схемах Индекса готовности к сетевому миру (рисунок 3) и Национального индекса развития цифровой экономики (рисунок 4).

1.3 Международные стандарты мониторинга развития информационного общества

Возрастающая роль ИКТ для социально-экономического развития способствовала повышенному вниманию ведущих международных организаций к гармонизации статистического наблюдения за развитием и использованием технологий в различных сферах деятельности на национальном уровне и, соответственно, к формированию международных стандартов в области мониторинга развития и использования цифровых технологий.

Основными разработчиками международных стандартов статистического наблюдения являются следующие организации (для каждой из них приводятся основные инструменты, в том числе и регулярно обновляемые).

1. Партнерство по измерению ИКТ для развития:

- список ключевых показателей и методических рекомендаций (ИКТ-инфраструктура, использование цифровых технологий домохозяйствами и населением, э-правительство, ИКТ в бизнесе, ИКТ в образовании) [40];
- показатели ИКТ для мониторинга Целей устойчивого развития ООН [41].

2. Организация экономического сотрудничества и развития (ОЭСР):

- определение и методология измерения электронной коммерции [42];
- модельное обследование бизнеса по вопросам использования ИКТ [7];
- модельное обследование населения/ домохозяйств по вопросам использования ИКТ [43];
- группировка видов экономической деятельности, относящейся к информационной индустрии (ИКТ-сектор и сектор контента и СМИ) [42];
- классификаторы товаров и услуг ИКТ [42];
- группировка кодов международной патентной классификации, относящихся к ИКТ [42];
- руководство по измерению ИКТ в здравоохранении [44].

3. Международный союз электросвязи (МСЭ):

- справочник по сбору административных данных в области электросвязи/ИКТ [45];
- руководство МСЭ по измерению доступа к ИКТ и их использования на уровне домохозяйств и отдельных лиц [46].

4. Институт статистики ЮНЕСКО:

- руководство по измерению использования ИКТ в образовании [47].

5. Евростат:

- руководство по обследованию использования ИКТ предприятиями и домохозяйствами [48].

6. Группа двадцати:

- инструментарий для измерения цифровой экономики [49];
- дорожная карта по созданию общего фреймворка для измерения цифровой экономики [50].

Признанным авторитетом и главным разработчиком международных стандартов измерения развития и использования ИКТ является ОЭСР, которая ведет эту деятельность с конца 1990-х. ОЭСР разработала стандарты измерения предметных областей мониторинга развития информационного общества, связанных с производством и использованием ИКТ, которые представлены в виде единой концептуальной схемы мониторинга [42], содержащей вопросы, на которые отвечают статистика и модельные обследования (рис. 5):

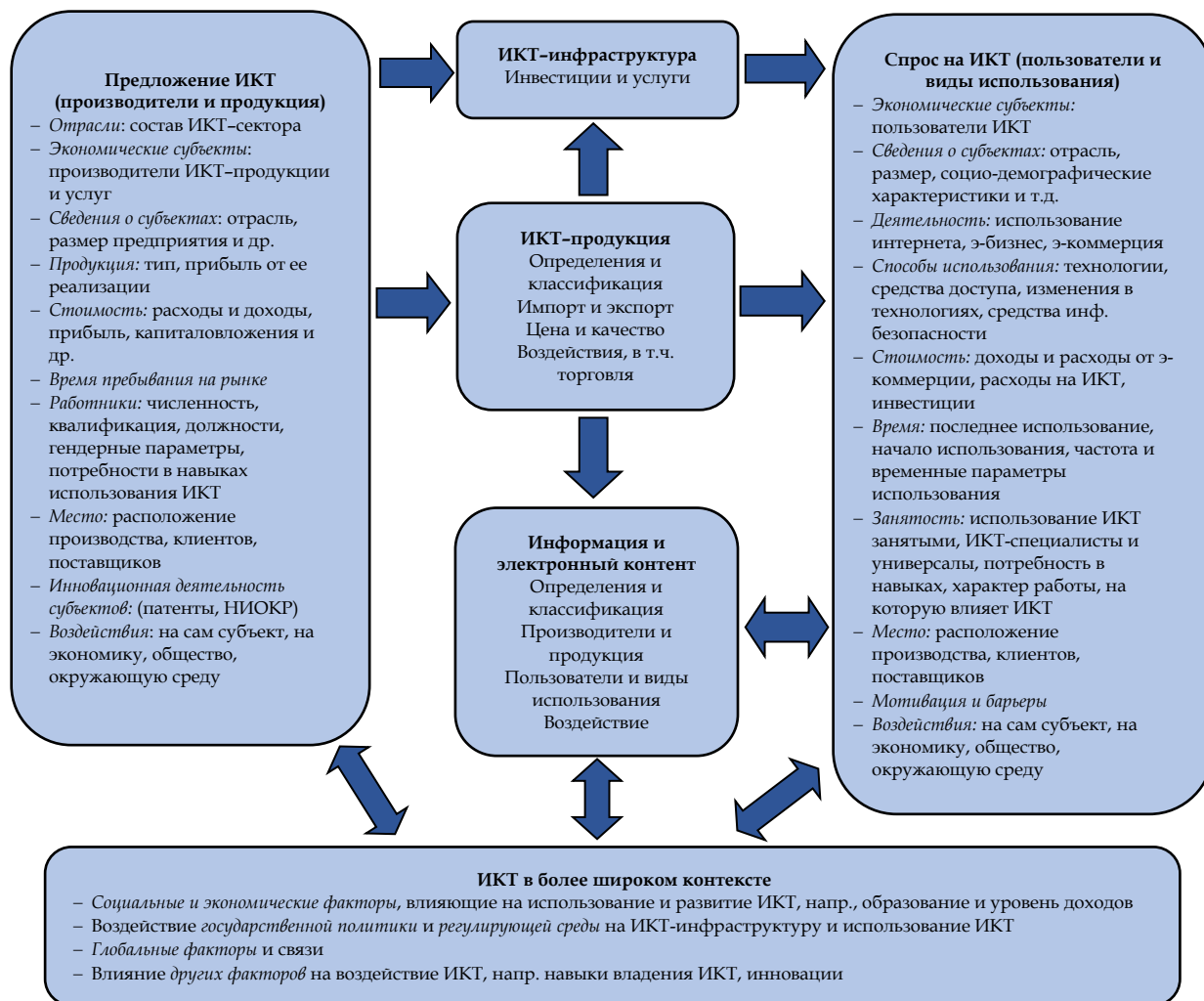


Рис. 5. Концептуальная схема мониторинга развития информационного общества ОЭСР
Источник: [42]

Концептуальная схема развития информационного общества включает в качестве основных компонентов производство ИКТ-товаров и услуг (поставщики – ИКТ сектор и сектор контента и СМИ), использование ИКТ бизнесом и населением, ИКТ-инфраструктуру, а также факторы, влияющие на развитие и использование ИКТ.

В целом состояние дел с разработкой международных стандартов мониторинга развития информационного общества обстоит следующим образом.

1. Наиболее проработаны (с точки зрения международных стандартов измерения) наблюдения за использованием ИКТ бизнесом и населением/домохозяйствами (сформированы и регулярно применяются модельные обследования и анкеты ОЭСР и Евростата).

Другие важные направления использования ИКТ (например, в государственном управлении, здравоохранении, образовании) со своей спецификой пока не имеют устоявшихся стандартов измерения и регулярной системы сбора данных для международных сопоставлений. В частности, разработанный ОЭСР подход к измерению цифрового здравоохранения [44] был однократно применен в 10 пилотных странах и не стал инструментом регулярного мониторинга на уровне

национальных статистических агентств, к тому же в нем не отражены некоторые современные тенденции цифровой трансформации здравоохранения.

Несмотря на то, что в общей концептуальной схеме измерения информационного общества ОЭСР (рисунок 5) обозначена предметная область мониторинга факторов, влияющих на производство, использование и воздействие ИКТ, она остается недостаточно проработанной с точки зрения формирования целостной системы показателей и модельного инструментария для их измерения. В частности, фактически не разработана методология измерения факторов, связанных с государственной политикой и нормативным регулированием.

2. Общая проблема у международных стандартов измерения предметных областей использования цифровых технологий – практически полное отсутствие или недостаточное внимание к новым цифровым технологиям третьей волны [1], которые стали драйверами современных процессов цифровой трансформации: виртуальная и дополненная реальность, искусственный интеллект, анализ больших данных, распределенный реестр (блокчейн), интернет вещей, робототехника и другие. Лишь в последние годы ОЭСР доработала модельные обследования бизнеса и населения/домохозяйств [7, 43], а Евростат в 2016–2020 гг. разработал анкеты [8-10], которые включают вопросы, использования облачных вычислений, социальных сетей, аналитики данных, искусственного интеллекта, цифровых платформ, робототехники, 3D-печати.

3. Методология измерения на макроуровне воздействия ИКТ остается неразработанной, включенные в модельное обследование ОЭСР показатели воздействия цифровых технологий на (микро)уровне предприятий пока не пилотировались, по ним нет данных Евростата и опыта измерения.

Тем не менее, можно утверждать, что основной тенденцией развития международных стандартов мониторинга развития и использования ИКТ на социально-экономическое развитие является все более полный учет показателей сложившейся триады областей оценки: готовность (факторы, в которые иногда включают производство цифровых товаров и услуг) – использование – воздействие.

1.4 Международные стандарты статистического наблюдения за использованием больших данных

Как отмечено в предыдущем разделе, в последнее время основные международные организации в области статистического наблюдения за развитием информационного общества стали модернизировать свои стандарты наблюдения для учета массового применения цифровых технологий третьей волны, в том числе – технологий работы с большими данными. В числе первых вопросы об использовании аналитики больших данных в свои модельные анкеты включили ОЭСР и Евростат. Их анализ приводится ниже.

1.4.1 Модельное обследование ОЭСР

В 2015 г. была опубликована новая редакция модельного обследования ОЭСР по вопросам использования ИКТ бизнесом [7], содержащая раздел об использовании аналитики больших данных и позволяющая собрать следующие сведения.

1. Использование аналитики [больших] данных.
2. Расходы на аналитику [больших] данных.
3. Мотивация для использования аналитики [больших] данных.
4. Воздействие аналитики [больших] данных.
5. Барьеры, ограничивающие или препятствующие использованию аналитики [больших] данных.

1.4.2 Обследование Евростата

Евростат оперативно отреагировал на модельное обследование ОЭСР и включил раздел об использовании аналитики больших данных в свои модельные анкеты для бизнеса 2016, 2018 и 2020 гг. [8-10].

Показатели в обследованиях Евростата отличались от разработанных ОЭСР: из пяти показателей в анкету 2016 г. был включен только один об использовании аналитики данных (проведение анализа больших данных собственными силами или сторонней организацией), но добавлен новый показатель об источниках данных для анализа.

В 2019 г. модельная анкета Евростата (для обследования предприятий в 2020 г.) была доработана, и в нее вошли следующие показатели использования аналитики больших данных на предприятии [10].

1. Используемые методы анализа больших данных с двумя вариантами ответов.
2. Наличие у предприятия какой-то компании или организации, проводившей для него анализ больших данных.
3. Рассмотрение предприятием возможности использования аналитики больших данных (для предприятий, которые не используют технологии работы с большими данными).
4. Барьеры, ограничивающие использование аналитики [больших] данных или препятствующие ему (модифицированный показатель ОЭСР).
5. Продажа предприятием больших данных (или доступа к ним) в отчетном году.
6. Покупка предприятием больших данных (или доступа к ним) в отчетном году.

Опубликованные на сайте Евростата данные обследования предприятий в 2020 г. показывают, что в среднем по ЕС анализ больших данных из разных источников (своими силами или по заказу) проводят 14% всех предприятий и более трети (34%) крупных предприятий с числом занятых больше 250 человек (рисунок 6).

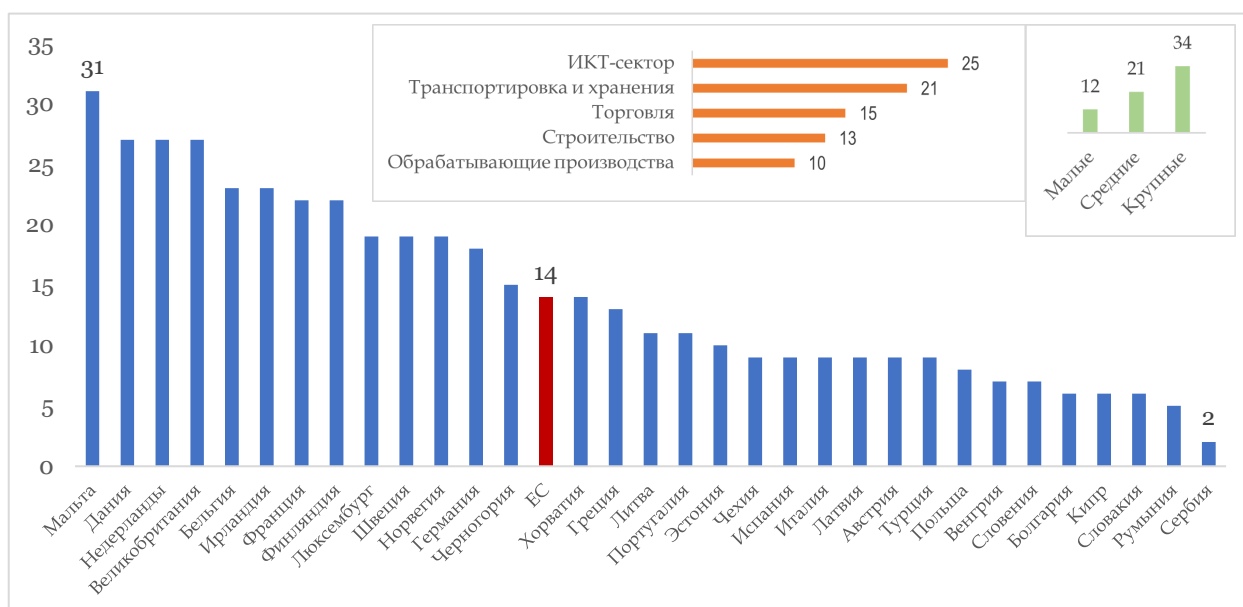


Рисунок 6. Доля предприятий в странах ЕС и ряда других стран, использующих анализ больших данных, 2020, %

Источник: [51]

Наиболее востребована аналитика больших данных в ЕС в таких отраслях как ИКТ-сектор (проводят анализ больших данных 25% предприятий), транспортировка и хранение (21%). Характерно, что в основном предприятия проводят анализ больших данных свои силами (полностью или частично, отдавая часть работы на аутсорсинг – 13% от всех предприятий). Среди трех представленных в анкете вариантов источников данных (сенсоры, геолокация, социальные сети) наиболее часто для анализа используются данные, генерируемые социальными сетями (7% в среднем по ЕС).

1.4.3 Выводы

Следует отметить, что в разработке международных стандартов статистического наблюдения за использованием технологий работы с большими данными сделаны только первые шаги. Технологии работы с большими данными рассматриваются как частный случай цифровых технологий и для них не выстроена комплексная система показателей, охватывающая производство, использование и воздействие этих технологий, а также факторов, влияющие на эти процессы. В

новом модельном обследовании ОЭСР сделана попытка сформулировать показатели воздействия технологий работы с большими данными на различные аспекты и результаты деятельности предприятий, однако эта часть методологии пока не реализована Евростатом.

Требует дальнейшего развития методология измерения расходов на разработку и использование технологий работы с большими данными, чтобы иметь возможность оценки объема рынков данных и др. вопросов. Необходимо расширение таксономии и направлений использования предприятиями других технологий работы с большими данными – пока выделены только машинное обучение, а также обработка естественного языка, генерация естественного языка или распознавание речи.

В целом можно сделать вывод, что за предложенными метриками и модельными вопросами не стоит проработанная концептуальная схема комплексного мониторинга развития, использования и воздействия технологий работы с большими данными на микро- и макроуровнях.

2 Концептуальная схема системы мониторинга BD4DE

Концептуальная схема мониторинга и оценки процессов производства, использования и воздействия технологий работы с большими данными представляет собой совокупность предметных областей, структура и взаимосвязи которых определяются: (а) целями мониторинга; (б) опытом применения цифровых технологий для развития отраслей экономики и секторов социальной сферы; (в) результатами научных исследований развития и использования цифровых технологий, включая технологии работы с большими данными; (г) международной и отечественной практикой разработки концептуальных схем оценки электронной готовности, мониторинга и оценки уровня развития цифровой экономики, моделей цифровой зрелости организаций и сфер деятельности.

Концептуальная схема BD4DE предоставляет основу для выбора и структурирования системы показателей для комплексного мониторинга процессов производства, использования и воздействия технологий работы с большими данными, обеспечивающих функционирование механизма обратной связи в системе управления цифровой экономикой.

Анализ современных научных исследований процессов технологического развития приводит к заключению, что потенциал развития и использования цифровых технологий для социально-экономического развития страны может быть реализован только при наличии адекватной государственной политики и регулирования, выстроенной системы управления и финансирования, кадров с необходимыми цифровыми компетенциями, эффективной системы исследований и разработок в этой сфере, безопасной и развитой цифровой инфраструктуры.

С учетом этих представлений и исходя из целей данной работы в концептуальную схему мониторинга процессов развития и использования технологий работы с большими данными включены следующие предметные области (рисунок 7):

- имеющиеся предпосылки (факторы), влияющие на производство, использование и воздействие технологий хранения и анализа больших данных (государственная политика и регулирование, институты управления и финансирования, цифровая инфраструктура, человеческий капитал, исследования и разработки, информационная безопасность);
- производство технологий работы с большими данными;
- использование технологий работы с большими данными для цифровой трансформации различных сфер деятельности (государственного и муниципального управления, бизнеса, здравоохранения, образования, науки);
- социально-экономические эффекты использования технологий работы с большими данными.



Рисунок 7. Концептуальная схема системы мониторинга и оценки процессов производства, развития и использования технологий работы с большими данными (BD4DE)

Для каждой из предметных областей мониторинга и оценки разработана концептуальная схема и система показателей, представленные в соответствующих статьях данного номера журнала [52]-[64].

3 Методы измерения показателей и источники данных

Как уже отмечалось, национальные системы статистического наблюдения процессов цифровой трансформации находятся на начальном этапе развития, в том числе – в Российской Федерации. Проведённая в 2020 г. модернизация формы федерального статистического наблюдения № 3-информ в части использования технологий работы с большими данными и другие доступные (коммерческие) источники данных не решают проблему комплексного мониторинга развития и использования технологий работы с большими данными. Это связано с тем, что в рамках действующего статистического наблюдения применяется ограниченное число показателей, связанных с большими данными, а широкий круг важных для характеристики развития и использования этих технологий показателей остаются недоступными – в частности, показатели, характеризующие НИОКР и подготовку кадров, спрос и использование технологий работы с большими данными в различных сферах деятельности, эффекты от использования больших данных.

Поэтому полноценная реализация комплексной системы мониторинга развития и использования технологий работы с большими данными в Российской Федерации должна опираться на собственную систему сбора данных для расчета показателей. В данном разделе представлены основные источники и методы получения данных для расчета показателей, включая:

- кабинетные исследования (заказ и сбор из открытых источников данных статистического наблюдения, оценка документов стратегического планирования и нормативного регулирования в сфере работы с большими данными по разработанным критериям и показателям);
- выборочные представительные опросы организаций (для расчета показателей использования и воздействия технологий работы с большими данными, которые не доступны в действующей статистике);
- экспертные опросы;
- инструментальные методы анализа публикационной и патентной активности для характеристики сферы НИОКР в области технологий работы с большими данными.

3.1 Кабинетные исследования

3.1.1 Сбор и обработка данных национальной статистики

Как отмечалось выше, в официальной статистике практически отсутствуют специализированные показатели, связанные с большими данными – они ограниченно представлены только в форме № 3-информ за 2020 г. Вместе с тем для оценки факторов, влияющих на развитие и использование технологий работы с большими данными (цифровая инфраструктура, информационная безопасность и др.), а также для сопоставлений могут использоваться имеющиеся данные текущего статистического наблюдения.

Сбор данных для системы мониторинга BD4DE опирается на следующие официальные открытые источники статистической информации Росстата и отраслевых министерств (частично данные заказываются в ГМЦ Росстата в необходимых разрезах):

- использование цифровых технологий населением и домохозяйствами (для оценки информационной инфраструктуры и информационной безопасности): результаты федерального статистического наблюдения по форме № 1-ИТ;
- использование цифровых технологий (включая технологии работы с большими данными) в бизнесе: результаты федерального статистического наблюдения по форме № 3-информ;
- использование цифровых технологий в бюджетной сфере: результаты федерального статистического наблюдения по форме №3-информ (для оценки уровня использования цифровых технологий в организациях здравоохранения, науки, высшего образования, государственного и муниципального управления), а также данные ведомственной статистики (Министерства науки и высшего образования, Министерства просвещения, Министерства здравоохранения) – для образования, науки и здравоохранения;
- человеческий капитал: результаты федерального статистического наблюдения по форме № 3-информ – в части обеспеченности организаций ИКТ-специалистами; результаты обследования населения Росстатом по вопросам занятости – в части наличия ИКТ-специалистов среди занятых в экономике; данные Министерства науки и высшего образования о подготовке специалистов в разрезе специальностей и направлений подготовки, связанных с цифровыми технологиями; результаты выборочного обследования по форме № 1-Т(проф) о численности и потребности организаций в работниках по профессиональным группам;
- цифровая инфраструктура: данные Министерства цифрового развития, связи и массовых коммуникаций относительно проникновения мобильного и стационарного ШПД и др., а также отраслевая статистика и данные по форме № 3-информ о скорости подключения организаций к интернету;
- цифровой сектор экономики: данные Росстата по форме № 3-информ.

3.1.2 Анализ и оценка документов

Показатели некоторых предметных областей мониторинга (прежде всего – государственной политики и регулирования, системы управления и механизмов финансирования) не могут быть измерены с использованием статистических данных, большая их часть основывается на оценках существующих документов – стратегий, программ, законов и других нормативных актов.

При проведении оценок и расчете значений качественных показателей, как правило, используются балльные оценки по порядковой шкале от 1 до 5, для которых разработаны критерии и алгоритмы получения целочисленных значений.

Чтобы избежать субъективности и неоднозначности оценок, во всех возможных случаях применяются критерии оценки по факту наличия или отсутствия в оцениваемых документах того или иного положения (например, соответствующего раздела стратегии или предмета регулирования) – чек-листы. Такой подход позволяет обеспечить прозрачность и наглядность самой процедуры получения значения показателей.

3.2 Опросы организаций предпринимательского сектора

Основным источником сведений для расчета показателей системы мониторинга BD4DE являются выборочные представительные опросы организаций.

При определении параметров опроса (генеральная совокупность объектов наблюдения, параметры представительной выборки) для системы мониторинга BD4DE учитывалась целевая

аудитория мониторинга, в частности, органы государственной власти, для которых важно получить сведения об использовании технологий работы с большими данными в приоритетных отраслях экономики и социальной сферы, которые определены в национальных целях развития [65].

Так как использование аналитики больших данных стало одной из предметных областей мониторинга в обновленных международных стандартах (см. выше раздел 1.2), то при разработке отечественного инструментария учитывались методические рекомендации ОЭСР [42] и Евростата [48], чтобы иметь возможность проведения международных сопоставлений с ситуацией в развитых странах. Детали методологии опроса предприятий представлены в указанных документах, в данном разделе приведено краткое описание.

3.2.1 Генеральная совокупность и представительная выборка

При формировании генеральной совокупности объектов наблюдения из различных сфер деятельности в качестве инструментов формирования генеральной совокупности и выбора организаций для обследования используются базы данных ЕГРЮЛ и СПАРК.

При отборе организаций используется случайная стратифицированная выборка. Для формирования страт, в рамках которых будет проводиться случайный отбор организаций, используются три переменные (страты): вид экономической деятельности; размер организации; субъект Российской Федерации.

Виды экономической деятельности. Состав и детализация отраслей экономики и секторов социальной сферы, которые будут охвачены обследованием организаций, выбирались так, чтобы: (1) иметь возможность мониторинга и оценки отраслей экономики и секторов социальной сферы, которые определены в национальных целях развития Российской Федерации до 2030 года; (2) обеспечить международную сопоставимость по отдельным показателям (прежде всего со странами-членами ОЭСР и ЕС).

Поэтому в генеральную совокупность включены организации, входящие в следующие категории и группировки кодов Общероссийского классификатора видов экономической деятельности (ОКВЭД-2) [66]:

- А Сельское, лесное хозяйство, охота, рыболовство и рыбоводство
- С Обрабатывающие производства
- F Строительство
- G Торговля оптовая и розничная, ремонт автотранспортных средств и мотоциклов
- H Транспортировка и хранение
- I Деятельность гостиниц и предприятий общественного питания
- J Деятельность в области информации и связи
- K Деятельность финансовая и страховая
- L Деятельность по операциям с недвижимым имуществом
- M Деятельность профессиональная, научная и техническая (69–74)
- N Деятельность административная и сопутствующие дополнительные услуги
- 95.1 Ремонт компьютеров и коммуникационного оборудования
- Цифровой сектор экономики (группировка)
- Топливо-энергетический комплекс (группировка)
- Жилищно-коммунальное хозяйство (группировка)
- Бизнес (группировка части приведенных выше отраслей в соответствии с рекомендациями ОЭСР [42]).

Размер организации. Развитие и использование технологий работы с большими данными зависит от численности сотрудников в организации, поэтому для системы мониторинга BD4DE используются следующие страты (с учетом рекомендаций ОЭСР и Евростата):

- малые организации – 10–49 работников;
- средние организации – 50–249 работников;
- крупные организации – 250 и больше работников.

Географический охват. В максимальном возможном варианте реализации методологии (при наличии необходимых ресурсов) строится представительная выборка по субъектам Российской Федерации (еще одна страта).

Результаты обследования будут взвешиваться в соответствии с общим количеством организаций в страте, поэтому результаты будут представительны как по указанным выше разрезам, так и по генеральной совокупности в целом.

3.2.2 Методика проведения опроса

Способ проведения. Опрос проводится (в зависимости от бюджета) методом личных интервью по стандартизированной анкете и/или с использованием электронных средств коммуникаций, в том числе путем заполнения респондентами онлайн-формы.

Целевой респондент. В качестве основного респондента выбирается лицо, принимающее решения и несущее основную ответственность за вопросы, связанные с развитием и использованием цифровых технологий в организации (руководитель по информационным технологиям, руководитель по работе с данными, руководитель цифровой трансформации, руководитель структурного подразделения, отвечающего за ИКТ). В небольших организациях респондентом выбирается руководитель верхнего уровня и/или владелец (непосредственно участвующий в деятельности). В любом случае респондентом не должно быть лицо, отвечающее только за бухгалтерский учет. По ряду вопросов в качестве респондента может выступать один из руководителей организации.

Отчетный период. Отчетным периодом является предыдущий календарный год (там, где это указано). В случаях, когда отчетный период не указан, респонденты должны относить ответы к текущему положению дел (на момент обследования).

3.3 Экспертные опросы

В отдельных предметных областях мониторинга (таких как государственная политика и регулирование, система управления и механизмы финансирования, производство технологий – в части оценки уровня российских разработок) присутствуют трудно формализуемые параметры наблюдения, которые невозможно измерить методами, используемыми в других ситуациях – статистическими, социологическими, инструментальными. Для получения значений показателей в подобных случаях используются экспертные опросы.

3.3.1 Формирование пула экспертов

Для проведения экспертного опроса необходимо на начальном этапе мониторинга каждой предметной области методом «снежного кома» сформировать пул экспертов: предложить каждому эксперту из первоначального списка (помимо заполнения анкеты) назвать 5 экспертов в данной предметной области. Формирование экспертного пула методом «снежного кома» завершается, когда предложенные имена начинают повторяться и число экспертов практически перестает расти.

Для исключения эффекта замыкания экспертов на группе специалистов, разделяющих близкие позиции и взгляды на предметную область, метод «снежного кома» используется в режиме нескольких «входов». Например, для показателей, которые должны учитывать позиции представителей различных заинтересованных сторон (бизнеса, власти, гражданского общества, научно-образовательного сообщества), важно сформировать сбалансированный состав экспертов, например, юристов, работающих в корпоративном, государственном, академическом и некоммерческом секторах. Метод нескольких «входов» также применяется в тех случаях, когда профессиональное сообщество разделено на группы специалистов, имеющих различные взгляды на предметную область.

Для проведения экспертных оценок из сформированного пула экспертов выбираются специалисты для решения конкретных задач заочного индивидуального опроса и групповой экспертизы. При этом используются как случайный отбор экспертов, так и методы самооценки и взаимной оценки. Само формирование пула экспертов методом «снежного кома» является одной из возможных реализаций метода взаимной оценки, позволяющее составить рейтинг экспертов по числу полученных рекомендаций.

3.3.2 Объем выборки экспертов

Для проведения заочного индивидуального опроса по стандартизованной анкете формируется группа экспертов количеством 20–50 человек, в зависимости от предметной области.

Для проведения групповой экспертизы формируются экспертные комиссии по предметным областям в составе 5–11 человек, состоящие (по возможности) из экспертов, имеющих наивысший рейтинг.

3.3.3 Опрос экспертов и согласование оценок

Для получения экспертных оценок в системе мониторинга BD4DE проводится экспертный опрос с использованием разработанной стандартизованной анкеты с вопросами закрытого типа (с предложением вариантов ответа). Экспертная оценка проводится в два этапа.

На первом этапе проводится независимый заочный опрос экспертов по стандартизованной анкете. Полученные оценки усредняются и выносятся на заседание экспертной комиссии, которое проводится в очном режиме.

На заседании комиссии обсуждаются полученные оценки и аргументация экспертов. Полученные оценки либо верифицируются, либо корректируются. В ходе заседания (которое при необходимости тоже может проходить в два этапа) ставится цель добиться консенсуса в оценке. Если в ходе заседания экспертной комиссии консенсус не достигнут и остались расхождения, сохраняется средняя оценка, полученная по итогам экспертного опроса.

3.4 Инструментальные методы измерения публикационной и патентной активности

Одной из предметных областей мониторинга работы с большими данными является сфера исследований и разработок, для мониторинга которой используются показатели публикационной и патентной активности.

Проблемой для измерения показателей исследований и разработок в области технологий работы с большими данными является выделение этой области с точки зрения имеющихся классификаторов. Действующие классификаторы (областей науки, патентов, публикаций в библиографических базах данных) достаточно «грубы», не отражают эту узкую область и не позволяют измерять библиометрические показатели на основе стандартных поисковых инструментов.

Эта проблема решается за счет разработки и использования ключевых слов, характеризующих данную область. С их использованием строится поисковый запрос, позволяющий формировать для измерения массив публикаций и патентов, которые относятся к технологиям работы с большими данными.

Для измерения показателей публикационной и патентной активности в рамках проекта был разработан набор ключевых слов и сформирован поисковый образ для выделения публикаций и патентов в области технологий работы с большими данными. Разработка проводилась в несколько этапов.

На первом этапе на основе консультаций с экспертами в области технологий работы с большими данными и анализа публикаций был сформирован исходный список ключевых слов, характеризующих данную область. Основными критериями отбора терминов для ключевых слов были:

- релевантность – термин используется в рамках области технологий работы с большими данными;
- полезность для целей данного исследования – термин расширяет искомый массив и позволяет идентифицировать публикации, которые не имеют в названии или других метаданных термина “big data”;
- отсутствие «шума» в результатах поиска – т.е. основная часть публикаций, содержащих предлагаемый термин в названии, ключевых словах или аннотации, должна относиться к указанной области исследований и разработок.

На втором этапе предложенный список был верифицирован в ходе опроса экспертов по работе с большими данными. Предложенный набор ключевых слов был подтвержден, наибольшее число баллов, помимо “big data”, набрали ключевые слова “Hadoop” и “MapReduce”.

В итоге, на основе разработанного списка ключевых слов был сформирован следующий поисковый образ для публикаций в области технологий работы с большими данными,

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

учитывающий различные варианты написания терминов.

Указанный поисковый образ использовался для формирования массива публикаций за 2016–2020 гг. Поиск проводился в поле «Тема» (TS – поиск по названиям, ключевым словам и аннотациям) в базе данных WoS, в аналогичном поле в базе данных Scopus, а также в текстовых полях патентной базы Patentscope.

3.4.1 Измерение публикационной активности

Научные публикации являются основным результатом научной деятельности (прежде всего в фундаментальной науке), поэтому изучение их количественных характеристик используется при решении широкого круга исследовательских и практических задач: от анализа структуры переднего фронта исследований до оценки результатов научной деятельности. В данном случае – для мониторинга состояния исследований и разработок в области технологий работы с большими данными.

В качестве измеряемого массива публикаций в системе мониторинга используются научные публикации, индексируемые в системе Web of Science (WoS). Выбор WoS для анализа публикационной активности российских ученых связан с тем, что она является наиболее авторитетной международной библиографической системой, для индексации, в которой тщательно отбираются рецензируемые журналы и другие научные издания. Важным является и то обстоятельство, что для анализа массива публикаций, представленных в WoS, разработаны аналитические инструменты InCites [67], позволяющие получать широкий спектр библиометрических показателей.

Использование пакета InCites наряду со стандартными возможностями поисковой системы WoS позволяет отслеживать как динамику публикационной активности российских ученых в области технологий работы с большими данными в различных разрезах (по субтехнологиям, квартилям журналов, типам публикаций), так и показатели «качества» публикационной активности, включая показатели воздействия (средняя цитируемость; цитируемость, взвешенная по предметной области), показатели международного сотрудничества и показатели связей научной деятельности с бизнесом.

В ходе разработки методов измерения было принято решение использовать для поиска не только названия публикаций и ключевые слова, как это часто делается, но и аннотации. Проведенные эксперименты и экспертная оценка поисковых выдач показали, что подавляющее число публикаций, имеющих в аннотациях искомые ключевые слова, относятся к области технологий работы с большими данными.

3.4.2 Измерение патентной активности

Патенты – это право интеллектуальной собственности, выданное уполномоченными органами изобретателям, позволяющее им использовать свои изобретения и эксплуатировать их в течение ограниченного периода времени (обычно 20 лет). Патенты выдаются частным лицам, фирмам или другим юридическим лицам при условии, что изобретение соответствует определенным критериям: оно должно быть новым, включать изобретательский уровень (то есть быть неочевидным) и быть пригодным для промышленного применения.

Патентная статистика используется для измерения ключевых показателей инновационной деятельности. Они используются для характеристики инновационной активности корпораций, академических организаций, отраслей или целых стран как индикатор коммерческого потенциала результатов исследований и разработок, приоритетов научно-технического развития, трансфера технологий, инновационных связей академического и коммерческого секторов и т.д. [68].

Для целей данной работы в мониторинг масштабов исследований и разработок в области технологий работы с большими данными были включены показатели российских патентных заявок, поданных в соответствии с договором о патентной кооперации (Patent Cooperation Treaty, PCT) и опубликованных в международных базах данных.

Поиск проводился в патентной базе данных Patentscope Всемирной организации интеллектуальной собственности [69]. Учитывая, что заявки по процедуре PCT могут подаваться на национальных языках участников договора (включая русский), для корректных международных сопоставлений поиск российских заявок осуществляется на русском и английском языках, а для сравнения выбираются англоязычные страны.

4 Нормативно-справочная информация для мониторинга

Любая методология мониторинга и оценки, претендующая на объективность, должна удовлетворять ряду общих взаимосвязанных требований, таких как:

- отчуждаемость от разработчиков и воспроизводимость результатов измерений;
- сопоставимость измерений, осуществляемых в различные периоды времени, а также с измерениями в аналогичных системах мониторинга (например, в других странах);
- однозначность и операциональность определения единиц и группировок для наблюдения.

Выполнение отмеченных требований в значительной степени обеспечивается использованием согласованной системы нормативно-справочной информации мониторинга, включающей контролируемые словари, справочники и классификаторы.

В качестве примера можно привести задачу мониторинга и оценки уровня развития цифрового сектора экономики. Для реализации указанных требований к подобной методологии необходимо: (1) иметь критерии и инструменты выделения организаций, относящихся к данному сектору экономики, (2) использовать или разработать стандартизированные методы измерения показателей. В данном случае это обеспечивается применением соответствующих классификаторов – группировки кодов экономической деятельности, относящихся к цифровому сектору экономики (ИКТ-сектор и сектор контента и СМИ) на основе ОКВЭД-2, которая была утверждена [70] Минкомсвязью России в 2015 г. и гармонизирована с группировкой предложенной ОЭСР [42], а также использованием показателей, основанных на федеральных статистических наблюдениях за деятельностью организаций (они в России также гармонизированы с международными стандартами). Такой подход дает возможность осуществлять мониторинг экономических показателей цифрового сектора экономики и сопоставлять результаты с показателями стран ОЭСР.

В данном разделе представлены классификаторы, нормативные справочники и контролируемые словари, используемые в системе мониторинга BD4DE.

4.1 Виды экономической деятельности

Для решения задач мониторинга (определение генеральной совокупности и выборки обследуемых организаций, выбор отраслей экономики, в разрезе которых будет проводиться обследование и др.) в системе мониторинга BD4DE также используется ОКВЭД-2, который гармонизирован с действующей второй редакцией Статистической классификации видов экономической деятельности в Европейском Союзе (NACE Rev. 2) [71].

Помимо стандартных видов экономической деятельности (отраслей), которые входят в разделы (первый уровень классификации) или коды высокого уровня (второй и третий уровень классификации) ОКВЭД-2 в системе мониторинга BD4DE используются специальные группировки видов экономической деятельности. Специальные группировки нужны для определения и операционализации относительно (1) новых отраслей экономики, которые не выделены в действующем классификаторе, такие как цифровой сектор экономики; (2) отраслей экономики, которые по сложившейся в Российской Федерации практике выделяются в системе государственного управления и выходят за рамки отраслей, представленных в классификаторе (примерами могут служить топливно-энергетический комплекс и жилищно-коммунальное хозяйство).

4.1.1 Группировка видов экономической деятельности, относящихся к цифровому сектору экономики

Впервые определение и группировку видов экономической деятельности, относящихся к ИКТ-сектору, предложила ОЭСР в 1998 г. Впоследствии группировка дорабатывалась и адаптировалась к новой версии классификатора, к ней была добавлена группировка сектора контента и СМИ [42]. В 2015 г. приказом Минкомсвязи России [70] были утверждены аналогичные группировки на основе ОКВЭД-2, согласованные с группировками ОЭСР.

Сектор ИКТ определяется как совокупность видов экономической деятельности, связанных с производством товаров и услуг, предназначенных для выполнения функции обработки информации и коммуникации с использованием электронных средств, включая передачу и отображение информации. В состав сектора ИКТ входят производство, оптовая торговля ИКТ-товарами и оказание ИКТ-услуг (телекоммуникационных услуг и услуг на основе информационных технологий).

Сектор контента и СМИ определяется как совокупность видов экономической деятельности, осуществляющих производство, издание и распространение контента (информации, продукции культурного назначения и продукции, предназначенной для развлекательных целей).

Отмеченные выше группировки, к сожалению, не формируют основу для мониторинга производства товаров и услуг, связанных с технологиями работы с большими данными, так как не обладают необходимой степенью детализации. Вместе с тем они используются в системе мониторинга BD4DE для определения «базы» отдельных показателей (таких как доля производства товаров и услуг, связанных с технологиями работы с большими данными, в цифровом секторе экономики).

4.1.2 Группировка видов экономической деятельности, относящихся к топливно-энергетическому комплексу

Виды экономической деятельности (по ОКВЭД-2), относящиеся к ТЭК, зафиксированы в соответствующем постановлении Правительства РФ в 2019 г. [72].

В группировку ТЭК включена добыча, транспортировка и переработка нефти, газа и угля, а также производство и распределение электроэнергии.

4.1.3 Группировка видов экономической деятельности, относящихся к жилищно-коммунальному хозяйству

Собирательная классификационная группировка видов экономической деятельности «жилищно-коммунальное хозяйство» на основе ОКВЭД-2 была утверждена приказом Министерства строительства и жилищно-коммунального хозяйства РФ в 2016 году [73].

В группировку включены виды деятельности, связанные с обеспечением электрической энергией, газом и паром; с водоснабжением, водоотведением, организацией сбора и утилизации отходов, деятельностью по ликвидации загрязнений; с обслуживанием и уборкой помещений, зданий и улиц; с управлением эксплуатацией жилого фонда.

4.2 Товары и услуги

В дополнение к группировкам видов экономической деятельности ОЭСР разработала группировку товаров и услуг, производимых в цифровом секторе экономики. Аналогичные собирательные классификационные группировки товаров и услуг на основе Общероссийского классификатора продукции по видам экономической деятельности (ОКПД-2), гармонизированного с международными, также были утверждены приказом [70] Минкомсвязи России в 2015 г.

Эти группировки используются в качестве базы для расчета показателей экспорта и импорта товаров и услуг, связанных с технологиями работы с большими данными (например, доля экспорта товаров и услуг, связанных с технологиями хранения и анализа больших данных от общего объема экспорта товаров и услуг цифрового сектора экономики и др.).

4.3 Патенты

ОЭСР в рамках методологии мониторинга развития информационного общества предложила группировку кодов Международной патентной классификации, которые относятся к ИКТ [42].

Данная группировка используется в системе мониторинга BD4DE для построения показателей патентной активности в области технологий работы с большими данными (например,

при определении доли патентов по технологиям работы с большими данными в общем числе ИКТ-патентов российских разработчиков). Сам массив патентов по технологиям работы с большими данными для анализа выделяется на основе ключевых слов и поисковых образов, по процедуре, описанной в разделе 3.4.

5 Инструментарий для измерения показателей мониторинга

Для измерения предложенной в рамках проекта системы показателей были разработаны необходимые инструменты, включая:

- критерии оценки документов (стратегий и программ) для кабинетных исследований (в рамках мониторинга государственной политики и регулирования [52]);
- анкеты для обследования организаций по вопросам зрелости работы с большими данными, использования технологий работы с большими данными и барьеров, мешающих использованию (для мониторинга использования технологий работы с большими данными в бизнесе, государственном и муниципальном управлении, здравоохранении, образовании и науке [60-63]);
- анкета для представительного опроса организаций цифрового сектора экономики, входящих в индустрию больших данных (для мониторинга индустрии работы с большими данными [59]);
- анкеты для экспертных опросов по вопросам оценки регулирования, механизмов управления и финансирования, социально-экономических эффектов развития и использования технологий работы с большими данными, а также уровня развития технологий [52, 53, 59, 64].

Заключение

Представленная методология мониторинга и оценки работы с большими данными разработана в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации». Система мониторинга BD4DE прошла пилотную апробацию и была уточнена в рамках первого этапа пилотной реализации, проведенной в 2020 г. (тестирование всех ее составляющих с привлечением экспертов-социологов и профессионалов в соответствующих предметных областях).

В данной статье содержатся результаты доработки системы мониторинга BD4DE на втором этапе: по нескольким предметным областям проведены экспертные опросы и исследования на ограниченной выборке респондентов, по части показателей проведены полномасштабные исследования и измерения. Результаты пилотажа (см. детали в [52]-[64]) в полной мере подтвердили применимость методологии и позволили провести некоторые усовершенствования. Доработанная система мониторинга развития и использования технологий работы с большими данными может быть использована для проведения ежегодных полномасштабных мониторинговых исследований.

Дальнейшее развитие системы мониторинга BD4DE будет проводиться по нескольким направлениям:

- доработка методологии (прежде всего системы показателей и инструментария) с учетом итогов и опыта проведения обследований на большом массиве организаций и представительной выборке экспертов;
- разработка таксономии технологий работы с большими данными;
- разработка цифровых инструментов для автоматизации и поддержки процессов мониторинга (онлайн-форм и цифровых инструментов для проведения опросов, обработки их результатов, публикации результатов мониторинга и др.).

Опыт разработки и реализации системы мониторинга развития и использования технологий работы с большими данными также может быть использован для организации мониторинга использования других сквозных технологий в цифровой экономике Российской Федерации.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций

Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Часть работы выполнена в рамках государственного задания КарИЦ РАН (№ 0185–2019–0095).

Литература

1. Katz, Raul. Social and economic impact of digital transformation on the economy. International Telecommunications Union, July 2017. 41 p.
2. T.V. Ershova and Yuri E. Hohlov, "Digital Transformation Framework: Monitoring of Large-Scale Socio-Economic Processes", in Management of large-scale system development: Proceedings of the 2018 Eleventh International Conference, MLSD 2018, Moscow, Russia, October 1-3, 2018. IEEE, 2018. 1-3 p. <https://doi.org/10.1109/MLSD.2018.8551765>
3. Всемирный банк. 2016 год. Доклад о мировом развитии 2016 «Цифровые дивиденды». Всемирный банк, Вашингтон, округ Колумбия.
4. T.V. Ershova, Yuri E. Hohlov and Sergei B. Shaposhnik, "Methodology for Digital Economy Development Assessment as a Tool for Managing the Digital Transformation Processes", in Management of large-scale system development: Proceedings of the 2018 Eleventh International Conference, MLSD 2018, Moscow, Russia, October 1-3, 2018. IEEE, 2018. 1-3 p. <https://doi.org/10.1109/MLSD.2018.8551846>
5. Всемирный банк. 2021 год. Доклад о мировом развитии 2021 «Данные для лучшей жизни». Всемирный банк, Вашингтон, округ Колумбия.
6. Приказ Росстата N 424 от 30.07.2020 "Об утверждении форм федерального статистического наблюдения для организации федерального статистического наблюдения за деятельностью в сфере образования, науки, инноваций и информационных технологий". Приложение. URL: https://gks.ru/bgd/free/b16_27/IssWWW.exe/Stg/d05/pril_424_1.doc
7. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (дата обращения 01.10.2021).
8. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2016. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/a39ae859-8a16-4306-8020-ae06d3df3c91/Questionnaire%20ENT%202016.pdf> (дата обращения 01.10.2021).
9. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2018. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/1fbef4a1-4c31-4b6a-afe8-19ee6d7e3b0f/ICT-Entr%202018%20-%20Model%20Questionnaire%20V%201.2.pdf> (дата обращения 01.10.2021).
10. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2020. General outline of the survey. URL: https://circabc.europa.eu/sd/a/47b2dcfa-2eb9-4cc4-9e98-b93a85406d67/MQ_2020 ICT_ENT.pdf (дата обращения 01.10.2021).
11. Danish Dada. e-Readiness for developing countries: moving the focus from the environment to the users // The Electronic Journal of Information Systems in Developing Countries (2006), vol. 27, N 6, p. 1-14.
12. E-readiness assessment: Who is Doing What and Where? 28 February 2005 (updated). URL: <http://bridges.org> (дата обращения 01.10.2021).
13. Hassan Alaaraj, Fatimah Wati Ibrahim An Overview and Classification of E-Readiness Assessment Models // International Journal of Scientific and Research Publications, Volume 4, Issue 12, December 2014.
14. Valentina Đurek, Zabok Prilaz, Nina Begičević Ređep Review on e-readiness assessment tools // Central European Conference on Information and Intelligent Systems. 2016.
15. Computer Systems Policy Project - CSPP (2000). Readiness Guide for Living in the Networked World. Washington, DC: CSPP.
16. Readiness for the Networked World: A Guide for Developing Countries. Cambridge, March 2002. Center for International Development at Harvard University, 22 p. URL: <https://cyber.harvard.edu/readinessguide/guide.pdf> (дата обращения 01.10.2021)
17. APEC E-Commerce Readiness Assessment Guide, 2000. URL: <https://www.apec.org/Publications/2000/12/APEC-ECommerce-Readiness-Assessment-Guide-2000> (дата обращения 01.10.2021).

18. Risk E-Business: Seizing the Opportunity of Global E-Readiness. Washington, DC: McConell International. 2000.
19. Готовность России к информационному обществу. Оценка возможностей и потребностей широкомасштабного использования информационно-коммуникационных технологий. Под редакцией Т.В. Ершовой. – М.: Издательство Института развития информационного общества, 2001. – 113 с.
20. Готовность России к информационному обществу. Оценка ключевых направлений и факторов электронного развития. Аналитический доклад / Под ред. С.Б. Шапошника. – М.: Институт развития информационного общества, 2004. – 252 с
21. N. Choucri, V. Maugis, S. Madnick, M. Siegel and S. Gillet, S. O'Donnel, M. Best, H. Zhu, F. Haghseta. Global e-readiness - for what? Paper 177. Massachusetts Institute of Technology Cambridge. May 2003.
22. Measuring the Information Society Report 2017. Volume 1 – Geneva: International Telecommunication Union, 2017. – xi + 155 p.
23. Measuring the Information Society Report 2017. Volume 2 – Geneva: International Telecommunication Union, 2017. – viii + 251 p.
24. United Nations E-Government Survey 2020. - UN, New York, 2020. - 323 p. URL: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020>
25. The Digital Economy and Society Index (DESI). URL: <https://ec.europa.eu/digital-single-market/en/desi> (дата обращения 01.10.2021).
26. International Digital Economy and Society Index 2018. – 79 p. <https://doi.org/10.2759/745483>
27. The Global Information Technology Report 2016. Innovating in the Digital Economy. World Economic Forum, Cornell University, INSEAD URL: http://www3.weforum.org/docs/GITR2016/GITR_2016_full%20report_final.pdf
28. Network Readiness Index 2020: Benchmarking the Future of the Network Economy. Portulance Institute, October 2020. – 317 p. URL: <https://networkreadinessindex.org/> (дата обращения 01.10.2021).
29. Индекс готовности регионов России к информационному обществу. 2004–2005 / Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. – М.: Институт развития информационного общества, 2005. 224 с.
30. Индекс готовности регионов России к информационному обществу. 2005–2006. – М.: Институт развития информационного общества, 2007. - 244 с.
31. Индекс готовности регионов России к информационному обществу. 2007–2008 / Под ред. Ю.Е. Хохлова и С.Б. Шапошника. – М.: Институт развития информационного общества, 2009. - 256 с.
32. Индекс готовности регионов России к информационному обществу 2008–2009 / Под ред. Ю.Е. Хохлова и С.Б. Шапошника. – М.: Институт развития информационного общества, 2010. – 296 с.
33. Индекс готовности регионов России к информационному обществу. 2009–2010 / Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. – М.: Институт развития информационного общества, 2011. - 360 с.
34. Индекс готовности регионов России к информационному обществу 2010–2011. Анализ информационного неравенства субъектов Российской Федерации / Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. М.: 2012. 462 с
35. Индекс готовности регионов России к информационному обществу 2013–2014. Под. Ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. Москва. 2015. 526 с.
36. Индекс готовности регионов России к информационному обществу. URL: <http://www.eRegion.ru> (дата обращения 01.10.2021).
37. Национальный индекс развития цифровой экономики: Пилотная реализация. Госкорпорация «Росатом». Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. 2018. – 92 с.
38. Методика оценки уровня развития информационного общества в субъектах Российской Федерации (Приложение № 1 к протоколу заседания Совета по региональной информатизации Правительственной комиссии по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 20 апреля 2016 г. №172пр). URL: <https://digital.gov.ru/uploaded/files/metodika-otsenki-urovnya-razvitiya-informatsionnogo-obschestva-v-subektah-rf-proekt.pdf> (дата обращения 01.10.2021).
39. Минкомсвязь представила рейтинг информатизации регионов-2017 // D-russia.ru. 13.10.2017. URL: <http://d-russia.ru/minkomsvyaz-predstavila-rejting-informatizatsii-regionov-2017.html> (дата обращения 01.10.2021).
40. CORE LIST OF ICT INDICATORS. March 2016 version. Partnership on Measuring ICT for Development. 2016. URL: <https://www.itu.int/en/ITU->

- [D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf](#) (дата обращения 01.10.2021).
41. Background note prepared by the Partnership on Measuring ICT for Development: Joint proposal of ICT indicators for the Sustainable Development Goal (SDG) indicator framework. Partnership on Measuring ICT for Development. 2015. URL: <https://www.itu.int/en/ITU-D/Statistics/Documents/intlcoop/partnership/Partnership-Background-note-on-ICT-indicator-proposal-for-Expert-Group.pdf> (дата обращения 01.10.2021).
 42. OECD (2011), OECD Guide to Measuring the Information Society 2011, OECD Publishing. <http://dx.doi.org/10.1787/10.1787/9789264113541-en> (дата обращения 01.10.2021)
 43. The OECD Model Survey on ICT Access and Usage by Households and Individuals. 2nd Revision. 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Access-Usage-Households-Individuals.pdf> (дата обращения 01.10.2021).
 44. Draft Guide to Measuring ICTs in the Health Sector. OECD. 2015. URL: <https://www.oecd.org/health/health-systems/Draft-oecd-guide-to-measuring-icts-in-the-health-sector.pdf> (дата обращения 01.10.2021).
 45. Справочник по сбору административных данных в области электросвязи/ИКТ URL: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITC_IND_HBK-2011-PDF-R.pdf (дата обращения 01.10.2021).
 46. Руководство по измерению доступа к ИКТ и их использования на уровне домашних хозяйств и отдельных лиц. ИТУ. 2014. URL: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITCMEAS-2014-PDF-R.pdf (дата обращения 01.10.2021).
 47. Guide to Measuring Information and Communication Technologies (ICT) in Education. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (дата обращения 01.10.2021).
 48. Methodological Manual for the surveys on ICT usage in enterprises and households. URL: <https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp>
 49. Toolkit for Measuring the Digital Economy. G20. Argentina 2018. URL: <https://www.oecd.org/g20/summits/buenos-aires/G20-Toolkit-for-measuring-digital-economy.pdf> (дата обращения 01.10.2021).
 50. A roadmap toward a common framework for measuring the Digital Economy. Report for the G20 Digital Economy Task Force. SAUDI ARABIA, 2020. OECD. 2020. URL: <https://www.oecd.org/sti/roadmap-toward-a-common-framework-for-measuring-the-digital-economy.pdf> (дата обращения 01.10.2021).
 51. Eurostat. Browse statistics by theme. Digital economy and society. Database. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (дата обращения 01.10.2021).
 52. Орлов С.В., Паджев В.В., Хохлов Ю.Е. Государственная политика и регулирование работы с большими данными // Информационное общество. 2021. № 4-5. С. 33–52. https://doi.org/10.52605/16059921_2021_04_33
 53. Елизаров А.М., Паджев В.В., Хохлов Ю.Е. Система управления и механизмы финансирования работы с большими данными // Информационное общество. 2021. № 4-5. С. 53–65. https://doi.org/10.52605/16059921_2021_04_53
 54. Шапошник С.Б., Янышпен А.В. Человеческий капитал для работы с большими данными в Российской Федерации // Информационное общество. 2021. № 4-5. С. 66–89. https://doi.org/10.52605/16059921_2021_04_66
 55. Хохлов Ю.Е., Шапошник С.Б. Исследования и разработки в области работы с большими данными // Информационное общество. 2021. № 4-5. С. 90–109. https://doi.org/10.52605/16059921_2021_04_90
 56. Малахов В.А., Хохлов Ю.Е., Шапошник С.Б. Использование технологий работы с большими данными в российской науке // Информационное общество. 2021. № 4-5. С. 200–219. https://doi.org/10.52605/16059921_2021_04_200
 57. Ершов П.С., Хохлов Ю.Е. Цифровая инфраструктура для работы с большими данными // Информационное общество. 2021. № 4-5. С. 110–131. https://doi.org/10.52605/16059921_2021_04_110
 58. Катин А.В., Хохлов Ю.Е. Доверие и безопасность работы с большими данными в России // Информационное общество. 2021. № 4-5. С. 315–333. https://doi.org/10.52605/16059921_2021_04_315

59. Малахов В.А., Хохлов Ю.Е., Шапошник С.Б. Индустрия работы с большими данными // Информационное общество. 2021. № 4–5. С. 278–299.
https://doi.org/10.52605/16059921_2021_04_278
60. Ершов П.С., Хохлов Ю.Е., Шапошник С.Б. Использование технологий работы с большими данными в российском бизнесе // Информационное общество. 2021. № 4–5. С. 300–314.
https://doi.org/10.52605/16059921_2021_04_300
61. Катин А.В., Хохлов Ю.Е. Использование технологий работы с большими данными в системе государственного управления России // Информационное общество. 2021. № 4–5. С. 150–165.
https://doi.org/10.52605/16059921_2021_04_150
62. Елизаров А.М., Хохлов Ю.Е. Мониторинг использования технологий работы с большими данными в российской образовательной системе // Информационное общество. 2021. № 4–5. С. 166–184. https://doi.org/10.52605/16059921_2021_04_166
63. Катин А.В., Хохлов Ю.Е. Использование технологий работы с большими данными в российском здравоохранении // Информационное общество. 2021. № 4–5. С. 185–199.
https://doi.org/10.52605/16059921_2021_04_185
64. Малахов В.А., Хохлов Ю.Е., Шапошник С.Б., Юревич М.А. Большие данные: социальные и экономические эффекты // Информационное общество. 2021. № 4–5. С. 132–149.
https://doi.org/10.52605/16059921_2021_04_149
65. Указ Президента Российской Федерации от 21 июля 2020 г. N 474 «О национальных целях развития Российской Федерации на период до 2030 года». URL:
<http://publication.pravo.gov.ru/Document/View/0001202007210012> (дата обращения 01.10.2021).
66. Общероссийский классификатор видов экономической деятельности (ОКВЭД 2). Введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 31 января 2014 г. N 14-ст URL: <https://docs.cntd.ru/document/1200110162> (дата обращения 01.10.2021).
67. InCites Help. URL: <https://incites.help.clarivate.com/Content/home.htm> (дата обращения 01.10.2021).
68. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. J Big Data 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
69. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (дата обращения 01.10.2021).
70. Приказ Министерства связи и массовых коммуникаций РФ от 7 декабря 2015 г. № 515 "Об утверждении собирательных классификационных группировок "Сектор информационно-коммуникационных технологий" и "Сектор контента и средств массовой информации". (Зарегистрирован 19.01.2016 № 40636) URL:
<http://publication.pravo.gov.ru/Document/View/0001201601220005> (дата обращения 01.10.2021)
71. Statistical Classification of Economic Activities in the European Community, Rev. 2. (NACE Rev. 2) URL:
https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNm=NACE_REV2&StrLanguageCode=EN&StrLayoutCode=HIERARCHI (дата обращения 01.10.2021)
72. Постановление Правительства РФ от 2 февраля 2019 г. № 76 “Об утверждении минимальных объемов добычи, производства, переработки, сжижения, обогащения, преобразования, хранения, передачи, распределения, транспортировки, поставок, перевалки, перегрузки, отгрузки, реализации энергетических ресурсов, продуктов их переработки, снабжения ими, осуществляемых юридическими лицами и индивидуальными предпринимателями для их отнесения к субъектам государственной информационной системы топливно-энергетического комплекса”. (Дата опубликования: 05.02.2019) URL:
<http://publication.pravo.gov.ru/Document/View/0001201902050010?index=1&rangeSize=1> (дата обращения 01.10.2021)
73. Приказ Министерства строительства и жилищно-коммунального хозяйства РФ от 27 апреля 2016 г. № 286/пр "Об утверждении собирательных классификационных группировок отрасли жилищно-коммунального хозяйства". URL:
<https://www.garant.ru/products/ipo/prime/doc/71346238> (дата обращения 01.10.2021).

METHODOLOGY FOR MONITORING BIG DATA TECHNOLOGIES DEVELOPMENT AND USE

Ershova, Tatiana Viktorovna

Candidate of economical sciences

Institute of the Information Society, general director

Research and analytical journal "Informacionnoe obshchestvo" (Information Society), editor-in-chief

Moscow, Russian Federation

tatiana.ershova@iis.ru

Hohlov, Yuri Evgenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor

Moscow, Russian Federation

yuri.hohlov@iis.ru

Shaposhnik, Sergei Borisovich

Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher

Petrozavodsk, Russian Federation

sergei.shaposhnik@gmail.com

Abstract

A framework of BD4DE (Big Data for Digital Economy) and a methodology for monitoring big data technologies use in the digital economy of the Russian Federation are proposed. The analysis of numerous frameworks for monitoring the development and use of digital technologies for socio-economic development (from e-readiness assessments and information society development assessments to digital maturity and digital economy assessments). The BD4DE framework covers the subject areas of monitoring: development and use of big data technologies and their impact on socio-economic development, as well as factors that affect them (government policy and regulation, governance and financing institutions, digital infrastructure, human capital, R&D, information security). Recommendations are given on a system of measurable indicators characterizing the proposed subject areas, sources are indicated.

Keywords

big data; big data technologies; monitoring the development and use of digital technologies; monitoring framework; Big Data for Digital Economy; BD4DE; monitoring and evaluation methodology; big data technologies development; use of big data; big data impact; enabling factors for big data development and use

References

1. Katz, Raul. Social and economic impact of digital transformation on the economy. International Telecommunications Union, July 2017. 41 p.
2. T.V. Ershova and Yuri E. Hohlov, "Digital Transformation Framework: Monitoring of Large-Scale Socio-Economic Processes", in Management of large-scale system development: Proceedings of the 2018 Eleventh International Conference, MLSD 2018, Moscow, Russia, October 1-3, 2018. IEEE, 2018. 1-3 p. <https://doi.org/10.1109/MLSD.2018.8551765>
3. Vsemirnyy bank. 2016 god. Doklad o mirovom razvitii 2016 «Tsifrovyye dividendy». Vsemirnyy bank, Vashington, okrug Kolumbiya.
4. T.V. Ershova, Yuri E. Hohlov and Sergei B. Shaposhnik, "Methodology for Digital Economy Development Assessment as a Tool for Managing the Digital Transformation Processes", in Management of large-scale system development: Proceedings of the 2018 Eleventh International Conference, MLSD 2018, Moscow, Russia, October 1-3, 2018. IEEE, 2018. 1-3 p. <https://doi.org/10.1109/MLSD.2018.8551846>
5. Vsemirnyy bank. 2021 god. Doklad o mirovom razvitii 2021 «Dannyye dlya luchshey zhizni». Vsemirnyy bank, Vashington, okrug Kolumbiya.

6. Prikaz Rosstata N 424 ot 30.07.2020 "Ob utverzhdenii form federal'nogo statisticheskogo nablyudeniya dlya organizatsii federal'nogo statisticheskogo nablyudeniya za deyatel'nost'yu v sfere obrazovaniya, nauki, innovatsiy i informatsionnykh tekhnologiy". Prilozheniye. URL: https://gks.ru/bgd/free/b16_27/IssWWW.exe/Stg/d05/pril_424_1.doc
7. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (accessed on 01.10.2021).
8. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2016. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/a39ae859-8a16-4306-8020-ae06d3df3c91/Questionnaire%20ENT%202016.pdf> (accessed on 01.10.2021).
9. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2018. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/1fbef4a1-4c31-4b6a-afe8-19ee6d7e3b0f/ICT-Entr%202018%20-%20Model%20Questionnaire%20V%201.2.pdf> (accessed on 01.10.2021).
10. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2020. General outline of the survey. URL: https://circabc.europa.eu/sd/a/47b2dcfa-2eb9-4cc4-9e98-b93a85406d67/MQ_2020_ICT_ENT.pdf (accessed on 01.10.2021).
11. Danish Dada. e-Readiness for developing countries: moving the focus from the environment to the users // The Electronic Journal of Information Systems in Developing Countries (2006), vol. 27, N 6, p. 1-14.
12. E-readiness assessment: Who is Doing What and Where? 28 February 2005 (updated). URL: <http://bridges.org> (accessed on 01.10.2021).
13. Hassan Alaaraj, Fatimah Wati Ibrahim An Overview and Classification of E-Readiness Assessment Models // International Journal of Scientific and Research Publications, Volume 4, Issue 12, December 2014.
14. Valentina Đurek, Zabok Prilaz, Nina Begičević Ređep Review on e-readiness assessment tools // Central European Conference on Information and Intelligent Systems. 2016.
15. Computer Systems Policy Project - CSPP (2000). Readiness Guide for Living in the Networked World. Washington, DC: CSPP.
16. Readiness for the Networked World: A Guide for Developing Countries. Cambridge, March 2002. Center for International Development at Harvard University, 22 p. URL: <https://cyber.harvard.edu/readinessguide/guide.pdf> (accessed on 01.10.2021).
17. APEC E-Commerce Readiness Assessment Guide, 2000. URL: <https://www.apec.org/Publications/2000/12/APEC-ECommerce-Readiness-Assessment-Guide-2000> (accessed on 01.10.2021).
18. Risk E-Business: Seizing the Opportunity of Global E-Readiness. Washington, DC: McConell International. 2000.
19. Gotovnost' Rossii k informatsionnomu obshchestvu. Otsenka vozmozhnostey i potrebnostey shirokomasshtabnogo ispol'zovaniya informatsionno-kommunikatsionnykh tekhnologiy. Pod redaktsiyey T.V. Yerzhovoy. – M.: Izdatel'stvo Instituta razvitiya informatsionnogo obshchestva, 2001. – 113 s.
20. Gotovnost' Rossii k informatsionnomu obshchestvu. Otsenka klyuchevykh napravleniy i faktorov elektronnoy razvitiya. Analiticheskiy doklad / Pod red. S.B. Shaposhnika. – M.: Institut razvitiya informatsionnogo obshchestva, 2004. – 252 s.
21. N. Choucri, V. Maugis, S. Madnick, M. Siegel and S. Gillet, S. O'Donnel, M. Best, H. Zhu, F. Haghseta. Global e-readiness - for what? Paper 177. Massachusetts Institute of Technology Cambridge. May 2003.
22. Measuring the Information Society Report 2017. Volume 1 – Geneva: International Telecommunication Union, 2017. – xi + 155 p.
23. Measuring the Information Society Report 2017. Volume 2 – Geneva: International Telecommunication Union, 2017. – viii + 251 p.
24. United Nations E-Government Survey 2020. - UN, New York, 2020. - 323 p. URL: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020>
25. The Digital Economy and Society Index (DESI). URL: <https://ec.europa.eu/digital-single-market/en/desi> (accessed on 01.10.2021).
26. International Digital Economy and Society Index 2018. – 79 p. <https://doi.org/10.2759/745483>

27. The Global Information Technology Report 2016. Innovating in the Digital Economy. World Economic Forum, Cornell University, INSEAD URL: http://www3.weforum.org/docs/GITR2016/GITR_2016_full%20report_final.pdf
28. Network Readiness Index 2020: Benchmarking the Future of the Network Economy. Portulance Institute, October 2020. – 317 p. URL: <https://networkreadinessindex.org/> (accessed on 01.10.2021).
29. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. 2004–2005 / Pod red. T.V. Yershovoy, Yu.Ye. Khokhlova, S.B. Shaposhnika. – M.: Institut razvitiya informatsionnogo obshchestva, 2005. - 224 s.
30. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. 2005–2006. – M.: Institut razvitiya informatsionnogo obshchestva, 2007. - 244 s.
31. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. 2007–2008 / Pod red. YU.Ye. Khokhlova i S.B. Shaposhnika. – M.: Institut razvitiya informatsionnogo obshchestva, 2009. - 256 s.
32. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu 2008–2009 / Pod red. YU.Ye. Khokhlova i S.B. Shaposhnika. – M.: Institut razvitiya informatsionnogo obshchestva, 2010. – 296 s.
33. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. 2009–2010 / Pod red. T.V. Yershovoy, YU.Ye. Khokhlova, S.B. Shaposhnika. – M.: Institut razvitiya informatsionnogo obshchestva, 2011. - 360 s.
34. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu 2010–2011. Analiz informatsionnogo neravenstva sub"yektov Rossiyskoy Federatsii / Pod red. T.V. Yershovoy, YU.Ye. Khokhlova, S.B. Shaposhnika. M.: 2012. 462 s
35. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu 2013–2014. Pod. Red. T.V. Yershovoy, YU.Ye. Khokhlova, S.B. Shaposhnika. Moskva. 2015. 526 s.
36. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. URL: <http://www.eRegion.ru> (accessed on 01.10.2021).
37. Natsional'nyy indeks razvitiya tsifrovoy ekonomiki: Pilotnaya realizatsiya. Goskorporatsiya «Rosatom». Pod red. T.V. Yershovoy, Yu.Ye. Khokhlova, S.B. Shaposhnika. 2018. – 92 s.
38. Metodika otsenki urovnya razvitiya informatsionnogo obshchestva v sub"yektakh Rossiyskoy Federatsii (Prilozheniye № 1 k protokolu zasedaniya Soveta po regional'noy informatizatsii Pravitel'stvennoy komissii po ispol'zovaniyu informatsionnykh tekhnologiy dlya uluchsheniya kachestva zhizni i usloviy vedeniya predprinimatel'skoy deyatel'nosti ot 20 aprelya 2016 g. №172pr). URL: <https://digital.gov.ru/uploaded/files/metodika-otsenki-urovnya-razvitiya-informatsionnogo-obshchestva-v-subektah-rf-proekt.pdf> (accessed on 01.10.2021).
39. Minkomsvyaz' predstavila reyting informatizatsii regionov-2017 // D-russia.ru. 13.10.2017. URL: <http://d-russia.ru/minkomsvyaz-predstavila-rejting-informatizatsii-regionov-2017.html> (accessed on 01.10.2021).
40. CORE LIST OF ICT INDICATORS. March 2016 version. Partnership on Measuring ICT for Development. 2016. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf (accessed on 01.10.2021).
41. Background note prepared by the Partnership on Measuring ICT for Development: Joint proposal of ICT indicators for the Sustainable Development Goal (SDG) indicator framework. Partnership on Measuring ICT for Development. 2015. URL: <https://www.itu.int/en/ITU-D/Statistics/Documents/intlcoop/partnership/Partnership-Background-note-on-ICT-indicator-proposal-for-Expert-Group.pdf> (accessed on 01.10.2021).
42. OECD (2011), OECD Guide to Measuring the Information Society 2011, OECD Publishing. <http://dx.doi.org/10.1787/10.1787/9789264113541-en> (accessed on 01.10.2021).
43. The OECD Model Survey on ICT Access and Usage by Households and Individuals. 2nd Revision. 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Access-Usage-Households-Individuals.pdf> (accessed on 01.10.2021).
44. Draft Guide to Measuring ICTs in the Health Sector. OECD. 2015. URL: <https://www.oecd.org/health/health-systems/Draft-oecd-guide-to-measuring-icts-in-the-health-sector.pdf> (accessed on 01.10.2021).

45. Spravochnik po sboru administrativnykh dannykh v oblasti elektrosvyazi/IKT URL: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITC_IND_HBK-2011-PDF-R.pdf (accessed on 01.10.2021).
46. Rukovodstvo po izmereniyu dostupa k IKT i ikh ispol'zovaniya na urovne domashnikh khozyaystv i otdel'nykh lits. ITU. 2014. URL: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITCMEAS-2014-PDF-R.pdf (accessed on 01.10.2021).
47. Guide to Measuring Information and Communication Technologies (ICT) in Education. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (accessed on 01.10.2021).
48. Methodological Manual for the surveys on ICT usage in enterprises and households. URL: <https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp>
49. Toolkit for Measuring the Digital Economy. G20. Argentina 2018. URL: <https://www.oecd.org/g20/summits/buenos-aires/G20-Toolkit-for-measuring-digital-economy.pdf> (accessed on 01.10.2021).
50. A roadmap toward a common framework for measuring the Digital Economy. Report for the G20 Digital Economy Task Force. SAUDI ARABIA, 2020. OECD. 2020. URL: <https://www.oecd.org/sti/roadmap-toward-a-common-framework-for-measuring-the-digital-economy.pdf> (accessed on 01.10.2021).
51. Eurostat. Browse statistics by theme. Digital economy and society. Database. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (accessed on 01.10.2021).
52. Orlov S.V., Padzhev V.V., Hohlov Yu.E. Gosudarstvennaya politika i regulirovaniye raboty s bol'shimi dannyimi // Informatsionnoye obshchestvo. 2021. № 4-5. S. 33-52. https://doi.org/10.52605/16059921_2021_04_33
53. Elizarov A.M., Padzhev V.V., Hohlov Yu.E. Sistema upravleniya i mekhanizmy finansirovaniya raboty s bol'shimi dannyimi // Informatsionnoye obshchestvo. 2021. № 4-5. S. 53-65. https://doi.org/10.52605/16059921_2021_04_53
54. Shaposhnik S.B., Yanyshen A.V. Kadry dlya raboty s bol'shimi dannyimi v Rossiyskoy Federatsii // Informatsionnoye obshchestvo. 2021. № 4-5. S. 66-89. https://doi.org/10.52605/16059921_2021_04_66
55. Hohlov Yu.E., Shaposhnik S.B. Issledovaniya i razrabotki v oblasti raboty s bol'shimi dannyimi // Informatsionnoye obshchestvo. 2021. № 4-5. S. 90-109. https://doi.org/10.52605/16059921_2021_04_90
56. Malakhov V.A., Hohlov Yu.E., Shaposhnik S.B. Ispol'zovaniye tekhnologiy raboty s bol'shimi dannyimi v rossiyskoy nauke // Informatsionnoye obshchestvo. 2021. № 4-5. S. 200-219. https://doi.org/10.52605/16059921_2021_04_200
57. Ershov P.S., Hohlov Yu.E. Tsifrovaya infrastruktura dlya raboty s bol'shimi dannyimi // Informatsionnoye obshchestvo. 2021. S. 110-131. https://doi.org/10.52605/16059921_2021_04_110
58. Katin A.V., Hohlov Yu.E. Doveriye i bezopasnost' raboty s bol'shimi dannyimi v Rossii // Informatsionnoye obshchestvo. 2021. S. 315-333. https://doi.org/10.52605/16059921_2021_04_315
59. Malakhov V.A., Hohlov Yu.E., Shaposhnik S.B. Industriya raboty s bol'shimi dannyimi // Informatsionnoye obshchestvo. 2021. S. 278-299. https://doi.org/10.52605/16059921_2021_04_278
60. Ershov P.S., Hohlov Yu.E. Ispol'zovaniye tekhnologiy raboty s bol'shimi dannyimi v rossiyskom biznese // Informatsionnoye obshchestvo. 2021. № 4-5. S. 300-314. https://doi.org/10.52605/16059921_2021_04_300
61. Katin A.V., Hohlov Yu.E. Ispol'zovaniye tekhnologiy raboty s bol'shimi dannyimi v sisteme gosudarstvennogo upravleniya Rossii // Informatsionnoye obshchestvo. 2021. № 4-5. S. 150-165. https://doi.org/10.52605/16059921_2021_04_150
62. Elizarov A.M., Hohlov Yu.E. Monitoring ispol'zovaniya tekhnologiy raboty s bol'shimi dannyimi v rossiyskoy obrazovatel'noy sisteme // Informatsionnoye obshchestvo. 2021. S. 166-184. https://doi.org/10.52605/16059921_2021_04_166
63. Katin A.V., Hohlov Yu.E. Ispol'zovaniye tekhnologiy raboty s bol'shimi dannyimi v rossiyskom zdavookhranenii // Informatsionnoye obshchestvo. 2021. № 4-5. S. 185-199. https://doi.org/10.52605/16059921_2021_04_185

64. Malakhov V.A., Hohlov Yu.E., Shaposhnik S.B., Yurevich M.A. Bol'shiye dannyye: sotsial'nyye i ekonomicheskiye efekty // Informatsionnoye obshchestvo. 2021. № 4–5. S. 132–149.
https://doi.org/10.52605/16059921_2021_04_149
65. Ukaz Prezidenta Rossiyskoy Federatsii ot 21 iyulya 2020 g. N 474 «O natsional'nykh tselyakh razvitiya Rossiyskoy Federatsii na period do 2030 goda». URL:
<http://publication.pravo.gov.ru/Document/View/0001202007210012> (accessed on 01.10.2021)
66. Obshcherossiyskiy klassifikator vidov ekonomicheskoy deyatel'nosti (OKVED 2). Vveden v deystviye prikazom Federal'nogo agentstva po tekhnicheskomu regulirovaniyu i metrologii ot 31 yanvarya 2014 g. N 14-st URL: <https://docs.cntd.ru/document/1200110162> (accessed on 01.10.2021).
67. InCites Help. URL: <https://incites.help.clarivate.com/Content/home.htm> (accessed on 01.10.2021)
68. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. J Big Data 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
69. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (accessed on 01.10.2021)
70. Prikaz Ministerstva svyazi i massovykh kommunikatsiy RF ot 7 dekabrya 2015 g. № 515 "Ob utverzhdenii sobiratel'nykh klassifikatsionnykh gruppirovok "Sektor informatsionno-kommunikatsionnykh tekhnologiy" i "Sektor kontenta i sredstv massovoy informatsii". (Zaregistrirovan 19.01.2016 № 40636) URL:
<http://publication.pravo.gov.ru/Document/View/0001201601220005> (accessed on 01.10.2021)
71. Statistical Classification of Economic Activities in the European Community, Rev. 2. (NACE Rev. 2) URL:
https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NACE_REV2&StrLanguageCode=EN&StrLayoutCode=HIERARCHI (accessed on 01.10.2021)
72. Postanovleniye Pravitel'stva RF ot 2 fevralya 2019 g. № 76 "Ob utverzhdenii minimal'nykh ob'yemov dobychi, proizvodstva, pererabotki, szhizheniya, obogashcheniya, preobrazovaniya, khraneniya, peredachi, raspredeleniya, transportirovki, postavok, perevalki, peregruzki, otgruzki, realizatsii energeticheskikh resursov, produktov ikh pererabotki, snabzheniya imi, osushchestvlyayemykh yuridicheskimi litsami i individual'nymi predprinimatel'yami dlya ikh otneseniya k sub'yektam gosudarstvennoy informatsionnoy sistemy toplivno-energeticheskogo kompleksa". (Data opublikovaniya: 05.02.2019) URL:
<http://publication.pravo.gov.ru/Document/View/0001201902050010?index=1&rangeSize=1> (accessed on 01.10.2021).
73. Prikaz Ministerstva stroitel'stva i zhilishchno-kommunal'nogo khozyaystva RF ot 27 aprelya 2016 g. № 286/pr "Ob utverzhdenii sobiratel'nykh klassifikatsionnykh gruppirovok otrasli zhilishchno-kommunal'nogo khozyaystva". URL:
<https://www.garant.ru/products/ipo/prime/doc/71346238> (accessed on 01.10.2021)

Информационное общество: политика и факторы развития

ГОСУДАРСТВЕННАЯ ПОЛИТИКА И РЕГУЛИРОВАНИЕ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Орлов Степан Владимирович

Кандидат экономических наук, доцент

Московский государственный университет имени М.В. Ломоносова, заведующий кафедрой истории общественных движений и политических партий

Москва, Российская Федерация

faculty@hist.msu.ru

Паджев Валентин Валентинович

Институт развития информационного общества (ИРИО), руководитель дирекции правовых программ

Москва, Российская Федерация

vpadzhev@iis.ru

Хохлов Юрий Евгеньевич

Кандидат физико-математических наук, доцент

Институт развития информационного общества, председатель совета директоров

Российский экономический университет имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО

Москва, Российская Федерация

yuri.hohlov@iis.ru

Аннотация

Представлены концептуальная схема и набор показателей, разработанные для мониторинга государственной политики и регулирования работы с большими данными как фактора, влияющего на развитие, производство, использование технологий работы с большими данными. В концептуальную схему включены показатели, характеризующие стратегическое планирование, нормативное правовое и нормативное техническое регулирование, саморегулирование и регулирование процессов обеспечения доверия и информационной безопасности при работе с большими данными. Предложены методы измерения показателей мониторинга государственной политики и регулирования работы с большими данными, приведены результаты измерения доступных показателей.

Ключевые слова

технологии работы с большими данными; государственная политика работы с большими данными; правовое регулирование работы с большими данными; техническое регулирование работы с большими данными; саморегулирование работы с большими данными; регулирование доверия и безопасности работы с большими данными; Big Data for Digital Economy; BD4DE

Введение

Развитие информационного общества на современном этапе связывают с процессами масштабного использования цифровых технологий в основных сферах жизнедеятельности, в том числе благодаря технологическим инновациям, базирующимся на работе с данными [1]. Исследования последних лет показали, что развитие цифровой экономики и достижение социальных и экономических эффектов от развития и использования цифровых технологий невозможны, если для этого не сформированы благоприятные условия, носящие отнюдь не технологический характер [2]. К ним относятся государственная политика и стратегическое планирование,

© Орлов С.В., Паджев В.В., Хохлов Ю.Е., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_33

нормативное регулирование, развитие человеческого капитала, исследования и разработки в области цифровых технологий и другие [3]. В настоящей работе исследуется влияние государственной политики и регулирования на использование технологий работы с большими данными в отдельно взятых организациях, на уровне отраслей экономики (или отдельных сфер деятельности) и в масштабах страны.

Под государственной политикой в данной статье понимается целенаправленная деятельность органов государственной власти по достижению и реализации общезначимых целей развития страны или отдельных сфер деятельности. В основе государственной политики лежат национальное законодательство, нормативное регулирование и правоприменительная практика, а ее реализация проявляется в процессах стратегического планирования, мерах регулятивного характера и финансовых приоритетах. Текущая государственная политика Российской Федерации до 2030 года зафиксирована в национальных целях развития, одной из которых является цифровая трансформация [4].

Как отмечено выше, цифровая трансформация экономики и общества базируется на применении как зрелых, так и нарождающихся сквозных цифровых технологий. К последним относятся технологии работы с большими данными, под которыми понимаются технологии сбора, обработки, хранения, анализа, представления, распространения, защиты, архивирования больших массивов данных. Поэтому регулярный мониторинг государственной политики в области работы с большими данными, осуществляемой в стране, регионе или отдельной сфере деятельности, позволит оценить влияние этого фактора и сформировать рекомендации по уточнению как политики, так и мер, направленных на ее реализацию.

В настоящей статье представлены компоненты концептуальной схемы мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации BD4DE (Big Data for Digital Economy) [5], относящиеся к государственной политике и нормативному регулированию работы с большими данными. На основе этой концептуальной схемы в 2020 году для отдельных показателей мониторинга получены оценки анализируемой сферы.

1 Определение предметной области мониторинга и оценки

В рамках проводимого исследования [5] под предметной областью мониторинга «Государственная политика и регулирование работы с большими данными» понимаются состояние и направления изменений государственной политики, общих правовых институтов, а также нормативного правового и технического регулирования работы с данными вообще и с большими данными в частности. В предметную область также включается нормативное регулирование процессов обеспечения доверия и безопасности работы с данными и морально-этическое регулирование (саморегулирование) использования данных и технологий работы с ними.

Государственная политика и регулирование работы с (большими) данными формируют благоприятную среду, устанавливают нормы и условия, определяющие принципы и границы использования технологий работы с данными. Это оказывает существенное влияние на развитие и использование соответствующих технологий и, в конечном счете, на экономику и социальную сферу.

Принципиальное отличие государственной политики и регулирования от других факторов состоит в том, что они непосредственно не воздействуют на развитие и использование технологий работы с данными и, как правило, являются реактивными, основанными на прецедентах, имеющихся барьерах и уже сформировавшихся потребностях.

Одним из первоочередных проявлений государственной политики в отдельной сфере деятельности является стратегическое целеполагание и планирование деятельности, направленной на достижение поставленных целей. В случае деятельности, связанной с развитием и использованием технологий работы с большими данными, следует оценить наличие и полноту соответствующих документов стратегического планирования. При этом важно провести сопоставительный анализ стратегического планирования процессов цифровой трансформации экономики и общества как в целом, так и применительно к работе с большими данными. Это связано с тем, что зачастую на уровне страны, региона или отдельной отрасли работу с данными рассматривают как составную часть процесса цифровой трансформации. В то же время исследования последних лет подтверждают первостепенную важность работы с данными [1] и

необходимость ее отдельной оценки и мониторинга, начиная со стадии стратегического планирования.

Еще одним проявлением государственной политики является нормативное правовое регулирование, выражающееся в наличии и полноте законов и подзаконных актов, устанавливающих правила работы с большими данными. Нормативное техническое регулирование в свою очередь характеризуется соответствием национальной системы стандартизации аналогичной международной системе, а также актуальностью и эффективностью стандартов, применяемых в стране. Развитие нормативного регулирования представляет собой достаточно медленный процесс, требующий согласования и учета интересов различных заинтересованных сторон. Он не всегда соответствует темпам цифрового развития и быстрой смены целых поколений технологий. Поэтому в последнее время все большее значение приобретают принципы саморегулирования (морально-этического регулирования) процессов работы с большими данными, вырабатываемые самими участниками.

Цифровая трансформация в разных сферах деятельности проходит неравномерно, поэтому оценку государственной политики и регулирования как факторов, влияющих на развитие и использование технологий работы с большими данными, необходимо анализировать и учитывать как на национальном уровне, так и в разрезе сфер деятельности – отраслей экономики, секторов социальной сферы и системы государственного управления.

2 Обзор подходов к разработке государственной политики и регулирования работы с большими данными

Большинство стран-лидеров развития цифровой экономики подчеркивает важность совершенствования инфраструктуры для работы с данными в масштабах страны или отдельной сферы деятельности. В недавнем докладе Организации экономического сотрудничества и развития (ОЭСР) «Индекс цифрового правительства 2019» [6] при оценке уровня цифровой зрелости системы государственного управления на национальном уровне отмечен многообещающий, но скромный прогресс в направлении использования цифровых технологий и данных для предоставления государственных услуг, ориентированных на пользователей. Авторами доклада сделан вывод, что во многих странах отсутствуют целенаправленная государственная политика и стратегический подход к развитию системы государственного управления, основанные на работе с данными. В то же время некоторые страны еще в начале 2010-х годов пересмотрели свои цифровые стратегии и приняли отдельные национальные стратегии в области работы с данными.

Правительство Соединенных Штатов Америки одним из первых поняло, что большие данные могут использоваться во всех сферах жизни, включая систему государственного управления. В 2012 г. началось финансирование Инициативы по исследованиям и разработкам технологий работы с большими данными [7], в которую изначально были вовлечены органы власти, бизнес, научно-исследовательские центры и университеты, а также некоммерческие организации. В 2014 г. была утверждена национальная стратегия работы с большими данными [8], в которой были зафиксированы основные положения государственной политики США по развитию и использованию больших данных граждан, бизнеса и органов власти, прежде всего для достижения экономических и социальных эффектов. В 2016 г. подкомитет Национального совета по науке и технологиям разработал Федеральный стратегический план по исследованиям и разработкам в области больших данных [9], включающий 7 направлений действий:

- создание возможностей следующего поколения путем использования новых оснований, методов и технологий работы с большими данными;
- поддержка исследований и разработок для изучения и понимания достоверности данных и извлекаемых из них знаний для принятия более эффективных решений, обеспечения прорывных открытий и уверенных действий;
- создание и совершенствование исследовательской информационной инфраструктуры, которая способствует внедрению инноваций на основе больших данных для выполнения функций ведомств;
- повышение ценности данных на основе политики, способствующей совместному использованию данных и управлению ими;
- понимание процессов сбора, обмена и использования больших данных с точки зрения конфиденциальности, безопасности и этики;

- улучшение национального образовательного ландшафта в области больших данных для удовлетворения растущего спроса как на талантливых аналитиков, так и на аналитические способности других работников;
- создание и расширение взаимосвязей в национальной инновационной экосистеме больших данных.

Дальнейшая эволюция государственной политики США в области работы с большими данными привела к тому, что в новой Федеральной стратегии «Использование данных как стратегического актива» [10], принятой в 2020 г., данные рассматриваются как национальный ресурс, имеющий огромную социальную ценность и используемый для решения задач в области здравоохранения, образования, национальной безопасности, энергетики. Стратегия и регулярно обновляемый план действий по ее реализации охватывают три направления, по которым сгруппированы правительственные инициативы (практики) в области стратегического управления работой с данными:

- формирование культуры, признающей ценность данных и способствующей их общественному использованию (практики 1–10);
- управление работой с данными, менеджмент и защита данных (практики 11–26);
- содействие эффективному и надлежащему использованию данных (практики 27–40).

Так, в текущем плане действий на 2020 г. содержится 20 мероприятий, «которые имеют важное значение для обеспечения процессов, наращивания потенциала и согласования усилий по более эффективному использованию данных в качестве стратегического актива», а также определен приоритет развития «зрелой среды управления активами данных» [11]. Мероприятия включают в себя определение потребностей пользователей, повышение квалификации персонала, оценку уровня зрелости инфраструктуры, укрепление структур управления, репозитории данных, этику данных и инструменты обработки данных.

К особенностям государственной политики США следует отнести подход, основанный на модели децентрализованного доступа к большим данным, в том числе путем привлечения негосударственных операторов для обеспечения функционирования государственных систем. При этом нормы, регламентирующие правомерный порядок использования таких данных, разбросаны по множеству различных актов. Федеральное законодательство США прямо не регулирует использование больших данных, оставляя это на усмотрение штатов и отдельных компаний, что является на сегодня достаточно большой редкостью с точки зрения нормативного регулирования использования больших данных. В то же время вопросам нормативного технического регулирования (стандартизации) работы с большими данными в США уделяется большое внимание. В 2013 г. Национальным институтом стандартов и технологий США (NIST) была начата активная деятельность по разработке и внедрению национальных стандартов в области больших данных. В составе NIST была сформирована рабочая группа по большим данным для стандартизации эталонной архитектуры больших данных, в которую вошли представители основных заинтересованных сторон – бизнеса, власти и научно-образовательного сообщества. В 2015 г. NIST впервые принимает серию стандартов в сфере терминологии, архитектуры больших данных, безопасности и конфиденциальности персональных данных при использовании соответствующих технологий [12]. Эти стандарты были пересмотрены и дополнены в 2018 [13] и 2019 [14] гг. В последующем национальные стандарты США были положены Международной организацией стандартизации и Международной электротехнической комиссией в основу серии международных стандартов в области больших данных (см. подробнее в [15]).

Целенаправленную политику в области работы с данными в качестве одного из основных направлений реализации своей цифровой стратегии проводит Европейский союз (ЕС), который в 2020 г. принял Европейскую стратегию в области данных [16], направленную на создание общего европейского пространства данных для функционирования единого рынка данных (со всего мира), где будут обеспечены:

- защита персональных и неперсональных данных (включая конфиденциальные данные бизнеса);
- межотраслевые потоки данных внутри ЕС;
- соблюдение европейских правил конфиденциальности, защиты и конкуренции;
- справедливые, практичные и понятные правила доступа к данным и их использования.

Достижение этой стратегической цели подразумевает:

- установление четких и справедливых правил доступа к данным и их повторного использования;
- инвестиции в следующее поколение стандартов, инструментов и инфраструктур для хранения и обработки данных;
- объединение усилий по наращиванию европейских облачных мощностей;
- объединение данных ключевых отраслей ЕС в общеевропейские и интероперабельные пространства данных;
- предоставление пользователям прав, инструментов и навыков для полного контроля своих данных.

В Европейской стратегии выделены четыре основных направления:

- обеспечение доступа к данным и их использования путем соблюдения требований интероперабельности при стратегическом управлении работой с данными;
- создание благоприятных условий для инвестиций в работу с данными, укрепление европейского потенциала и инфраструктуры для хранения, обработки и использования данных, интероперабельность;
- повышение компетенций при работе с данными, инвестирование в навыки работы с данными граждан, малых и средних предприятий;
- формирование общеевропейских пространств данных в стратегических отраслях экономики и сферах деятельности, имеющих значение для общества.

Определены девять приоритетных общеевропейских пространств данных: пространство промышленных (производственных) данных; пространство данных «зеленой повестки»; пространство данных мобильности; пространство данных здравоохранения; пространство финансовых данных; пространство данных энергетики; пространство сельскохозяйственных данных; пространство данных государственного управления; пространство данных образования, а также Европейское открытое научное облако.

Принятие данной стратегии демонстрирует высокую политическую значимость работы с данными для Европейского союза и его стремление стать основным игроком в глобальной экономике данных.

Основопологающим документом в области нормативного регулирования работы с персональными данными в ЕС является действующий с 2018 г. Общеевропейский регламент по защите персональных данных (далее – Регламент, General Data Protection Regulation, GDPR) [17], который предписывает всем членам Евросоюза гармонизировать с ним национальное законодательство. В названном Регламенте большие данные не выделены в отдельную категорию, но в нем значительно расширено понятие персональных данных, и в него включен ряд новых принципов, регламентирующих обработку и управление данными. Это принципы: прозрачности обработки данных; обеспечения приватности по умолчанию; обеспечения подотчетности операторов за соблюдение требований законодательства о персональных данных; обеспечения безопасности обрабатываемых данных и несколько других. В подходе Евросоюза к регулированию использования данных такое регулирование распространяется как на первичный сбор данных, так и на их обработку, обезличивание, использование и уничтожение. На большие данные, содержащие персональные данные, распространяются принципы обработки данных, закрепленные в Регламенте. Существенно, что Регламент согласован с положениями целого ряда международных правовых актов и, по существу, устанавливает законные основания для обработки данных.

Отметим, что в отличие от американской децентрализованной модели Евросоюз предусматривает создание Европейского совета по защите данных (European Data Protection Board), который обеспечивает централизованный контроль применения GDPR и единообразие политики государств-членов ЕС. В целом же в Евросоюзе планируется развитие рынка в первую очередь промышленных данных путем создания компаний-посредников, которые будут обеспечивать безопасный доступ, в том числе к государственным большим данным [17].

Следуя общей политике Евросоюза по работе с большими данными, государства-члены ЕС также разрабатывают и проводят собственные национальные политики. Так, в январе 2021 г. Германия утвердила Стратегию в области работы с данными, ставящую целью лидерство в инновационном и совместном использовании данных в Европе [18]. Ключевым приоритетом является более эффективное использование «инновационного потенциала», присущего данным, с

одновременным усилением соблюдения европейских ценностей, общих идей защиты данных и их суверенитета, превращением их в глобальную модель.

В стратегии заданы четыре основные направления, содержащие в общей сложности 240 мер для реализации государственной политики:

- разработка эффективной устойчивой инфраструктуры для работы с данными;
- расширение инновационного и ответственного использования данных;
- повышение компетенций в области работы с данными и формирование культуры данных;
- обеспечение органам власти роли первопроходца.

Меры государственной политики, направленные на развитие навыков, инфраструктуры и регулирования в сфере защиты персональных данных и платформ данных, а также обеспечение ведущей роли государственного сектора в экономике данных, считаются критическими для достижения политической цели превращения Германии в «пионера цифровых инноваций».

Отдельного рассмотрения заслуживает государственная политика Великобритании, являющейся одним из лидеров в развитии и использовании ИКТ в системе государственного управления, в том числе с точки зрения применения технологий работы с большими данными. Политика использования данных изначально велась в рамках движения открытых данных, прежде всего, открытых государственных данных. В 2013 году в Великобритании была принята первая версия «Стратегии использования возможности передачи данных», в которой, как и во многих других развитых странах, закреплялся принцип коллективности работы с данными – к этому процессу приглашались НКО, исследователи, представители бизнеса и другие заинтересованные стороны [19].

Сегодня работа с большими данными в Великобритании остается в центре внимания. В начале сентября 2021 г. правительство Великобритании запланировало финансирование в 2022–2026 гг. государственных проектов по хранению больших данных и аналитике для обеспечения ими департаментов центрального правительства и остального государственного сектора. Сама Национальная стратегия Великобритании в области данных [20] достаточно долгое время находится в стадии становления и доработки, проект основного стратегического документа был впервые опубликован на правительственном сайте gov.uk в 2019 г., вторая редакция – в декабре 2020 г. Последняя редакция Национальной стратегии от 18 мая 2021 г. содержит в качестве основы четыре тесно взаимосвязанные области эффективного использования данных: базис для работы с данными (например, связанный с повышением качества данных), навыки работы с данными; доступность данных и ответственное использование данных. В стратегии выделены пять основных направлений (миссий), предполагающих, в том числе, использование технологий работы с большими данными:

- 1) раскрытие ценности данных в масштабе всей экономики;
- 2) обеспечение нормативного режима работы с данными, способствующего росту и обеспечивающего доверие;
- 3) трансформация использования государственных данных для повышения эффективности и совершенствования государственных услуг;
- 4) обеспечение безопасности и устойчивости инфраструктуры для работы с данными;
- 5) возглавить международный обмен данными.

Декларируется [21], что движение по этим направлениям обеспечит британскому обществу социально-экономические выгоды от работы с данными, выражающиеся в экономическом росте, создании рабочих мест, улучшении качества государственных услуг и усилении эффективности научных исследований и инноваций, основанных на работе с данными.

Китайский подход к государственной политике и регулированию работы с большими данными основан на применении режима национального благоприятствования, когда основные требования по защите информации позволяют китайскому бизнесу и органам власти внедрять отечественные инновационные сервисы и системы.

В марте 2014 г. большие данные впервые прозвучали в отчете о работе правительства Китайской народной республики, в октябре 2015 г. было официально заявлено об осуществлении государственной стратегии работы с большими данными [22], а в 2016 г. была принята Национальная стратегия работы с большими данными. Китайская «Белая книга больших данных» 2019 г. [22] ознаменовала эволюцию к «глубокой интеграции больших данных и реальной экономики» и рассмотрению данных как стратегического актива, который будет способствовать

развитию экономики, улучшать общественное управление, повышать качество работы правительства, контроля и управления. Основной целью становится упор на ценность данных и их совместное использование, а также обеспечение качества и безопасности данных, что подразумевает наличие в Китае достаточно зрелой технической основы для функционирования систем работы с большими данными. Одновременно Китай надеется, что реализация принятой стратегии поможет развитию исследований и применению ключевых технологий, комплексному реформированию конкурентных преимуществ страны, а также позволит использовать решающий момент для будущего развития.

Заявляя амбициозные цели, руководство Китая, тем не менее, отмечало недостаток имеющихся отечественных цифровых решений для работы с данными, низкий уровень совместного использования и переиспользования накопленных массивов данных, недостаточный уровень обеспечения безопасности и защиты данных, в том числе персональных. В связи с этим был разработан и вступил в силу 1 сентября 2021 г. Закон о безопасности данных [23], призванный содействовать «развитию и использованию данных ... для упорядоченной и свободной» экономики данных. Сфера применения закона определена в статье 5: «Государство защищает права граждан и организаций, связанные с данными; поощряет законное, разумное и эффективное использование данных; обеспечивает законный и упорядоченный свободный поток данных и стимулирует развитие экономики данных, где данные рассматриваются как ключевой фактор для повышения благосостояния людей».

Несмотря на то, что принятый закон является нормативным актом, в нем зафиксированы основные направления развития экономики данных, такие как совместное управление с децентрализованными обязанностями; использование данных; стратегия работы с большими данными и развитие инфраструктуры для работы с данными, поддерживающие их инновационное использование в отраслях; фундаментальные исследования в области технологий обработки данных; создание системы стандартов в области данных; поддержку образования и профессиональной подготовки в области работы с данными; системы классификации данных, реестры и каталоги данных; роли и обязанности по обеспечению безопасности данных; прикладные исследования и разработки в области новых технологий обработки данных в целях получения экономических и социальных эффектов; посреднические услуги в области данных; электронное правительство; международная торговля данными; открытые платформы данных.

Принятие подобного всеобъемлющего закона закрепляет государственную политику Китайской народной республики в области работы с данными и определяет ключевые элементы, необходимые для реализации национальной стратегии: навыки, финансирование, нормативное регулирование, инфраструктура, использование данных. Китай рассматривает стать ключевым игроком в глобальной экономике данных и имеет четкий план действий, необходимых для достижения цели стать «сильной производственной державой, сильной кибердержавой и сильной цифровой державой».

В большинстве стран мира приняты соответствующие стратегии работы с данными и планы действий по их реализации, содержащие направления и мероприятия по нормативному регулированию. Анализ действующего законодательства зарубежных стран и моделей нормативного регулирования в области работы с данными можно найти в недавнем отчете Аналитического центра при Правительстве Российской Федерации [24].

Остановимся на подходах к государственной политике использования и регулирования больших данных в России. Российские органы власти, политики и представители бизнеса стали говорить о больших данных относительно недавно. Если понятие «открытые данные» в нормативном поле появилось с принятием «майских указов» Президента России в 2012 г., то понятие «большие данные» хотя и определено государственным стандартом [25], толкуется в России по-разному. Сегодня при обсуждении заинтересованными сторонами государственной политики в области больших данных фигурируют три основных подхода.

Первый подход заключается в том, что большие данные зачастую представляются огромными массивами сведений о пользователях различных сервисов (социальных сетей, банковских сервисов, мобильной связи и т. д.), поэтому только власть способна регулировать распоряжение этими сведениями и обязана выполнять функцию сохранения неприкосновенности персональных данных граждан, обеспечивать их безопасность. В рамках такого подхода предлагается создать государственный оператор больших данных, который собирал бы, хранил, обрабатывал и контролировал большие данные по всей стране. Однако существенной проблемой реализации

подобного подхода, как и использования больших данных российскими властями, является отсутствие качественных государственных массивов данных: примеров по-настоящему больших государственных данных в России немного – это данные Федеральной налоговой службы, Росреестра, Ростуризма и еще нескольких ведомств.

Второй подход предполагает, что власть будет выстраивать политику в области больших данных совместно с бизнесом и научно-образовательным сообществом, учитывая интересы всех заинтересованных сторон. Аналогичный подход использовался в России при разработке методических рекомендаций по работе с открытыми данными.

Третий подход ориентирован на крупных владельцев больших данных, которые не хотят, чтобы власть вмешивалась в сферу их деятельности, полагая, что любые ограничительные меры неминуемо ударят по бизнесу.

Все перечисленные подходы имеют свои преимущества и недостатки, выбор их оптимального объединения по-прежнему остается открытым.

Анализ, приведенный выше, свидетельствует, что факторы, связанные с государственной политикой и регулированием, существенно влияют на развитие и использование цифровых технологий и, в частности, технологий работы с большими данными, при этом в большинстве стран инициативы в области работы с большими данными тесно увязаны с нормативным регулированием в отношении персональных данных.

3 Концептуальная схема и система показателей

3.1 Концептуальная схема предметной области

Результаты анализа, представленные в предыдущем разделе, позволяют описать концептуальную схему мониторинга государственной политики и регулирования работы с большими данными в виде двух компонентов, охватывающих аспекты целеполагания и стратегического планирования, а также актуальности и полноты регулирования этих процессов (см. рисунок 1).

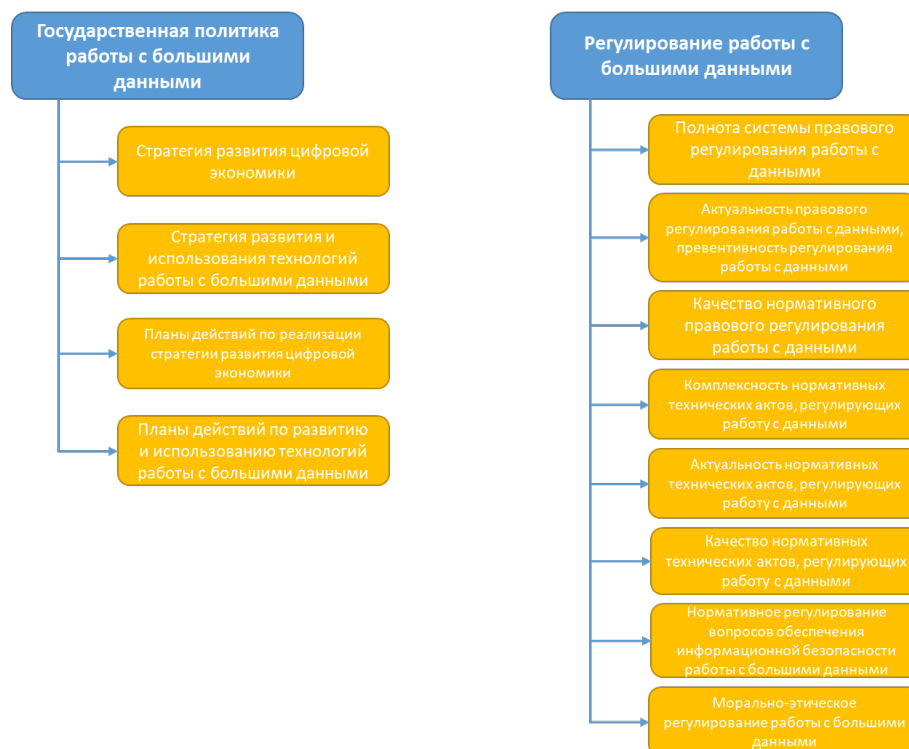


Рис. 1 - Концептуальная схема предметной области мониторинга «Государственная политика и регулирование работы с большими данным»

Для оценки целеполагания и уровня стратегического планирования используются показатели, характеризующие наличие и полноту соответствующих документов для

характеристики процессов цифровой трансформации (экономики и общества) на основе использования сквозных цифровых технологий.

На фоне наличия общих стратегических документов цифрового развития важно также оценить наличие в стране (регионе, отрасли) стратегических документов развития и использования технологий работы с большими данными, их комплексность и актуальность с точки зрения целей и задач работы с большими данными.

Для оценки реализуемости государственной политики, зафиксированной в принятых стратегиях, необходима оценка планов (программ) действий по реализации этих стратегий, так как их отсутствие, неполнота или необеспеченность ресурсами могут привести к дисбалансу и низкой эффективности расходования средств, потере доверия граждан и бизнеса, снижению благосостояния населения. Поэтому в систему показателей мониторинга анализируемой предметной области включены показатели, характеризующие наличие конкретных планов действий по реализации стратегических документов развития цифровой экономики, а также их реализуемость и обеспеченность ресурсами (например, финансовыми, кадровыми, научно-методическими).

Аналогичные показатели применяются и при оценке планов действий по реализации стратегии развития и использования технологий больших данных.

Нормативное правовое регулирование процессов работы с большими данными характеризуется показателями, которые позволяют оценить полноту, актуальность и качество законов и подзаконных актов, устанавливающих правила работы с большими данными. Полнота и актуальность национальной системы нормативного технического регулирования в свою очередь характеризуется ее соответствием международной системе стандартизации, а также действенностью (обязательностью, эффективностью) национальных и международных стандартов, принятых на территории страны. Как отмечалось ранее, в современных условиях все большее значение приобретают вопросы морально-этического регулирования (саморегулирования) процессов работы с большими данными, поэтому для характеристики этого фактора используется показатель наличия в стране (или отдельной сфере деятельности) подобных инициатив, выраженных в официальных действиях самих участников процесса (см., например, [6]).

При оценке действующей государственной политики и системы регулирования важно учитывать не только ситуацию в стране в целом, но и наиболее важные направления работы с большими данными, приводящие к максимальным экономическим и социальным эффектам. Поэтому большинство показателей данной предметной области мониторинга также рассматривается в разрезе отдельных сфер деятельности (отраслей экономики, секторов социальной сферы, системы государственного управления).

3.2 Государственная политика работы с большими данными

Ключевым показателем государственной политики в области цифрового развития является наличие долгосрочной национальной стратегии, имеющей в качестве основной цели широкомасштабное использование цифровых технологий и данных для социально-экономического развития:

- (ГПБД-01-01) *Стратегия развития цифровой экономики.*

Использование предложенного показателя позволяет охарактеризовать наличие и полноту политики, проводимой в масштабах всей страны, охватывающей все сферы деятельности и различные цифровые технологии, а не только технологии работы с большими данными. Подобная стратегия, как отмечалось в разделе 2, может иметь разные наименования (развития информационного общества, электронного развития, цифрового развития, цифровой трансформации и т.д.) и утверждаться либо как отдельный документ стратегического планирования, либо быть составной частью более общей национальной стратегии социально-экономического развития. Неравномерность процессов цифрового развития может привести к тому, что подобные стратегии могут приниматься как в масштабах всей страны, так и для отдельных сфер деятельности.

Наряду с общим показателем наличия стратегии развития цифровой экономики для целей настоящего исследования представляется необходимым использование показателя

- (ГПБД-01-02) *Стратегия развития и использования технологий работы с большими данными,*

который характеризует наличие (или отсутствие) долгосрочной стратегии работы с данными как на национальном уровне, так и в разрезе отдельных сфер деятельности. Подобная стратегия также может приниматься как отдельный документ стратегического планирования или быть составной частью более общей цифровой стратегии (или стратегического документа с названием, близким по смыслу).

Не менее важным документом стратегического планирования является план (программа) действий по реализации цифровых стратегий как для социально-экономического развития, так и для производства и использования технологий работы с большими данными. Для этого в системе мониторинга BD4DE предусмотрены следующие показатели:

- (ГПБД-01-03) *План действий по реализации стратегии развития цифровой экономики;*
- (ГПБД-01-04) *План действий по реализации стратегии развития и использования технологий работы с большими данными.*

Соответствующие документы стратегического планирования могут называться по-разному (программы, планы мероприятий, дорожные карты, проекты и т. д.), но по своей сути должны быть направлены на реализацию государственной политики в области цифрового развития и отражать реализуемость и обеспеченность ресурсами (например, финансовыми, кадровыми, научно-методическими) утвержденных цифровых стратегий. Приведенные показатели применяются как для страны в целом, так и для отдельных сфер деятельности.

3.3 Регулирование работы с большими данными

Развитие цифровых технологий оказывает существенное влияние на общественные отношения, что проявилось в новых формах взаимоотношений между субъектами права, осуществляемых с использованием цифровых технологий, замещении (или представлении) субъектов права кибернетическими устройствами, сосуществовании реального и виртуального пространств, деградации института географических и геополитических границ.

Полнота, качество и актуальность нормативного регулирования цифровой среды в значительной степени определяют возможность эффективного развития цифровой экономики, а также обеспечивают предупреждение негативных последствий использования цифровых технологий, в том числе технологий работы с большими данными. Именно эти факторы и будут оцениваться при мониторинге системы регулирования цифровой среды.

3.3.1 Правовое регулирование

Полнота нормативного правового регулирования работы с большими данными характеризуется определенностью понятийного аппарата в области работы с данными (субъекты правоотношений, состав их прав, условия доступа или ограничения доступа к данным), урегулированностью правоотношений в части повторного (совместного) использования данных, юридической значимости данных, открытости (доступности) данных. Этим целям служит показатель

- (ПРБД-02-01) *Полнота правового регулирования работы с данными,*

который оценивается как для страны в целом, так и для отдельных сфер деятельности.

Еще одной важной характеристикой нормативного правового регулирования может служить показатель

- (ПРБД-02-02) *Качество правового регулирования работы с данными,*

оценивающий единообразие регулирования в разрезе отдельных сфер деятельности, между ними, а также по уровням нормативного правового регулирования (федеральный, региональный, муниципальный). Кроме того, оценка качества нормативного регулирования должна обеспечивать возможность измерения степени исполнения и контроля исполнения нормативных правовых актов, регулирующих работу с данными; избыточность регулирования правоотношений, возникающих при работе с данными; наличие ответственности за несоблюдение норм, устанавливаемых законодательством в области работы с данными.

Мониторинг актуальности правового регулирования работы с данными будет проводиться с помощью показателя

- (ПРБД-02-03) *Актуальность правового регулирования работы с данными,*

который будет характеризовать регулирование виртуальных (кибернетических) субъектов правоотношений, связанных с данными, географических и геополитических границ в виртуальном

пространстве, а также его независимость от вида пространства (реального и виртуального), используемого для гражданско-правовых или иных отношений. Еще одним критерием актуальности национального или ведомственного правового регулирования работы с данными может служить степень соответствия национального законодательства международному.

3.3.2 Техническое регулирование

Для развития и использования технологий работы с большими данными техническое регулирование, направленное на обеспечение совместимости, сопоставимости и качества данных, значит не меньше, чем правовое регулирование. Как и в случае с нормативным правовым регулированием, в системе мониторинге BD4DE оцениваются полнота, качество и актуальность технического регулирования цифровой среды:

- (ТРБД-02-01) *Полнота технического регулирования работы с данными.*

Полнота действующего корпуса нормативных технических актов, регулирующих работу с большими данными, характеризуется определенностью в части терминологии области больших данных, построения и использования эталонной архитектуры больших данных и других вопросах:

- (ТРБД-02-02) *Качество технического регулирования работы с данными.*

Этот показатель также должен отражать единообразный подход к стандартизации работы с большими данными как в отдельных сферах деятельности, так и между ними. Кроме того, оценка качества зависит от возможности исполнения и контроля исполнения нормативных технических актов, избыточности технического регулирования, наличия ответственности за несоблюдение стандартов и иных нормативных технических актов.

Для оценки актуальности нормативного технического регулирования в области больших данных будет использован показатель

- (ТРБД-02-03) *Актуальность технического регулирования работы с данными,*

оценивающий степень соответствия и/или запаздывания системы национальных стандартов по отношению к системе международных стандартов в области больших данных.

3.3.3 Саморегулирование

Быстрое развитие цифровых технологий приводит к тому, что нормативное регулирование не всегда успевает оперативно реагировать на происходящие изменения, поэтому на первый план выходят нормы морально-этического регулирования работы с данными:

- (СРБД-02-01) *Морально-этическое регулирование работы с данными.*

Эти нормы характеризуются наличием принципов профессионального и этического поведения бизнеса, обеспечивающих соблюдение прав граждан и юридических лиц, являющихся субъектами и пользователями данных, снижение рисков, связанных с неосведомленностью граждан об угрозах вмешательства в частную жизнь, поддержание доверия общественности к деятельности по обработке, использованию и защите различных видов данных, обеспечение баланса интересов участников рынка данных, формирование и распространение добросовестных практик, а также содействие развитию цифрового сектора экономики. Зачастую эти принципы кодифицируют в виде того или иного публичного документа, которому обязуются следовать основные заинтересованные стороны.

3.3.4 Нормативное регулирование обеспечения доверия и безопасности

Проведенный анализ (см. раздел 2) показал, что особое место в нормативном регулировании занимают вопросы обеспечения доверия и безопасности при работе с данными. Для мониторинга и оценки в системе BD4DE будет использован показатель

- (ДББД-02-01) *Нормативное регулирование обеспечения доверия и информационной безопасности при работе с большими данными,*

отражающий наличие (или отсутствие) утвержденных требований к безопасности, актуальности и полноте массивов данных, а также обязательность применения этих требований и ответственность за их несоблюдение. Еще одной характеристикой нормативного регулирования обеспечения доверия и безопасности работы с большими данными может служить степень соответствия национального законодательства международному.

4 Методология исследования

Предметные области системы мониторинга BD4DE, связанные с государственной политикой и регулированием, не могут быть изучены и охарактеризованы с помощью статистических данных, практически все показатели носят качественный характер и связаны с анализом существующих документов – стратегий, программ, законов и других нормативных актов.

Основной использованный метод – это экспертные исследования и расчеты значений показателей с помощью балльных оценок по порядковой шкале от 1 до 5, для которых разработаны критерии и алгоритмы получения целочисленных значений. Для снижения уровня субъективности и неоднозначности экспертных оценок во всех возможных случаях применяются критерии наличия или отсутствия в оцениваемых документах отдельных положений (например, соответствующего целеполагания в стратегии или предмета нормативного регулирования).

Для получения значений показателей в трудно формализуемых параметрах мониторинга, которые невозможно измерить статистическими, социологическими или инструментальными методами, используются методы проведения экспертных опросов. Детальная процедура формирования экспертного пула, объема выборки, опроса экспертов и согласования экспертных оценок изложена в работе [5], посвященной разработке методологии мониторинга развития и использования технологий работы с большими данными.

5 Пилотное исследование и его результаты

5.1 Пилотное исследование методики оценки

Целями пилотного исследования были апробация и, при необходимости, корректировка методики оценки, в том числе уточнение методов оценки отдельных показателей, источников информации и инструментария, используемых для мониторинга государственной политики и регулирования технологий работы с большими данными.

Для проведения пилотного исследования по результатам первой стадии пилотной реализации предложенной концептуальной схемы мониторинга государственной политики и регулирования технологий работы с большими данными были разработаны анкеты для опроса экспертов и организаций.

Разработанные анкеты проанализированы экспертами в данной предметной области (специалистами по государственному управлению и регулированию) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций).

Экспертами по государственному управлению и регулированию были сделаны предложения по уточнению формулировок для обеспечения однозначного понимания сформулированных вопросов и учета специфики данной предметной области. Экспертом-социологом был сформулирован ряд предложений по уточнению и разъяснению для респондентов отдельных положений анкеты.

На последней стадии пилотной реализации полученные замечания и предложения экспертов были учтены в финальной версии анкеты для опроса экспертов и организаций, а также была произведена оценка фактора «Государственная политика и регулирование работы с большими данными» для Российской Федерации.

5.2 Пилотное исследование на основе предложенной методики

Пилотный мониторинг государственной политики и регулирования проводился в соответствии с концептуальной схемой, описанной в разделе 3.

Оценка документов стратегического планирования на национальном уровне, относящихся к развитию цифровой экономики и работе с большими данными – показатели (ГПБД-01-01) и (ГПБД-01-02), – заключалась, прежде всего, в анализе утвержденной стратегии развития информационного общества в Российской Федерации на 2017–2030 годы [26], где *развитие цифровой экономики* явно обозначено в качестве одной из целей. К недостаткам этой стратегии следует отнести отсутствие в документе критериев достижения целей и выполнения задач, зафиксированных в виде измеримых показателей мониторинга, которые должно было разработать Правительство Российской Федерации. Но до настоящего времени такие критерии так и не сформулированы.

Подготовка стратегии социально-экономического развития Российской Федерации до 2030 года, начатая по инициативе Правительства Российской Федерации в феврале 2021 г., должна была определить пути достижения национальных целей развития, поставленных в июльском указе Президента России [4]. Одной из национальных целей развития обозначена *цифровая трансформация* экономики и общества, которая вошла как составная часть в общую стратегию социально-экономического развития. Работа над стратегией завершилась утверждением в октябре 2021 г. Перечня инициатив социально-экономического развития Российской Федерации до 2030 года [27], в котором к национальной цели развития «Цифровая трансформация» отнесены 5 инициатив. Следует отметить, что стратегия развития цифровой экономики в Российской Федерации не ограничивается только целью «Цифровая трансформация»; для достижения других национальных целей развития предусмотрены дополнительные стратегические инициативы, непосредственно связанные с использованием цифровых технологий и данных для достижения целей социально-экономического развития. С каждой из стратегических инициатив связаны ключевые результаты, которые должны быть достигнуты к 2024 и 2030 годам, хотя далеко не со всеми ожидаемыми результатами связаны конкретные критерии и измеримые показатели достижения целей и выполнения задач.

Наряду с разработкой общей (фронтальной) стратегии велась работа и над стратегическими планированием цифровой трансформации отдельных сфер деятельности. В соответствии с национальными целями развития были определены приоритетные сферы деятельности:

- промышленность;
- сельское хозяйство;
- строительство;
- развитие городской среды;
- транспорт и логистика;
- энергетическая инфраструктура;
- финансовые услуги;
- здравоохранение;
- образование;
- наука;
- государственное и муниципальное управление;
- социальная сфера;
- экология и природопользование,

для которых были разработаны стратегические направления (отраслевые стратегии) цифровой трансформации. Состав и содержание стратегических направлений были рассмотрены и согласованы президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности [28], после чего они утверждались соответствующими распоряжениями Правительства Российской Федерации (см., например, [29]). По каждому стратегическому направлению определены приоритеты, цели и задачи, с которыми связаны измеримые показатели цифровой трансформации сферы деятельности. В то же время за пределами выделенных приоритетных сфер деятельности остаются отрасли экономики и сектора социальной сферы, которые в не меньшей степени нуждаются в цифровой трансформации.

Подводя итог проведенного кабинетного исследования, можно утверждать, что в Российской Федерации государственная политика развития цифровой экономики зафиксирована в нескольких действующих документах, однако уровень стратегического планирования в целом можно *оценить как удовлетворительный*.

Для оценки планов реализации стратегий (цифровой экономики, работы с большими данными – показатели (ГПБД-01-03) и (ГПБД-01-04) – проводился анализ планов на соответствие стратегиям и друг другу, наличие в них необходимых видов обеспечения (финансовое; кадровое; нормативное правовое и нормативное техническое, научно-методическое; организационное; инфраструктурное), их реализуемость и комплексность (охват приоритетных направлений и всех взаимосвязанных процессов). Под планом реализации при этом понимался утвержденный документ стратегического планирования, который мог иметь разные наименования (программа, план мероприятий, дорожная карта, проекты и т. д.), но должен предназначаться для достижения целей и решения задач соответствующих стратегий.

При оценке планов реализации стратегий развития цифровой экономики и стратегии работы с большими данными в Российской Федерации анализировались такие документы, как государственная программа «Информационное общество» [30], национальная программа «Цифровая экономика» [31] и входящие в нее федеральные проекты, ведомственные программы цифровой трансформации [32], программы цифровой трансформации государственных корпораций и организаций с государственным участием и т. д. Все эти документы в той или иной степени включают отдельные мероприятия, направленные на использование аналитики больших данных и применение технологий искусственного интеллекта (в первую очередь – машинного обучения), что, в свою очередь, неизбежно влияет на развитие технологий работы с большими данными и формирование самих массивов данных для обработки.

Анализ показал, что подавляющее большинство планов (программ) реализации стратегий как для развития цифровой экономики, так и для работы с большими данными демонстрирует обособленность процессов применения цифровых технологий в различных сферах деятельности или организациях. Это проявляется, в том числе, в отсутствии взаимоувязанности мероприятий, создании разрозненных информационных систем с несопоставимыми или трудно сопоставимыми массивами данных. Налицо отсутствие единой технической политики, представленной в форме обязательных для применения стандартов формирования моделей данных, управления основными (мастер-) данными, методами разметки слабоструктурированных данных и т. д., углубляющее неразбериху при работе с большими данными. Сложившийся подход ведет к накоплению больших массивов данных, но очень слабо обеспечивает рост ценности данных, к тому же приводит к трудностям при их совместном использовании и переиспользовании.

В целом проведенное кабинетное исследование показывает, что в Российской Федерации подходы к реализации государственной политики по работе с большими данными зафиксированы в нескольких действующих документах, однако планы реализации соответствующих документов стратегического планирования не могут быть оценены выше, чем на «удовлетворительно».

Мониторинг нормативного правового регулирования (показатели ПРБД-02-01, ПРБД-02-02 и ПРБД-02-03) включает анализ понятийного аппарата, состава и прав субъектов, для которых урегулированы правоотношения при работе с данными, условий доступа к данным. Дополнительными критериями оценки системы правового регулирования выступают единообразие подходов к регулированию в отдельных сферах деятельности и на различных уровнях системы государственного управления (федеральном, региональном, муниципальном), возможность контроля исполнения нормативных правовых актов, ответственность за их неисполнение, соответствие международному законодательству.

Частично требуемый анализ по состоянию на начало 2020 г. был проведен в докладе Аналитического центра при Правительстве Российской Федерации [24], и результаты настоящего кабинетного исследования лишь подтвердили выводы, сделанные в этом докладе. Дополнительное кабинетное исследование установило пробелы нормативного правового регулирования работы с данными практически по всем направлениям, начиная от неопределенности терминов «данные», «большие данные», «виртуальные» субъекты правоотношений и заканчивая неотделимостью данных от информационных систем, в которых они обрабатываются. Это приводит к проблемам с регулированием взаимоотношений при работе с данными (кроме персональных данных и данных, составляющих государственную тайну), которые подменяются регулированием взаимоотношений, связанных с разработкой и использованием информационных систем. Актуальность нормативного правового регулирования в Российской Федерации также оставляет желать лучшего – на фоне действующих нормативных правовых актов Евросоюза или Китая в области работы с большими данными (см. выше раздел 2) отставание составляет несколько лет.

Подводя итог, можно оценить ситуацию с нормативным правовым регулированием работы с данными в Российской Федерации как *неудовлетворительную*.

Национальное техническое регулирование работы с большими данными, которое характеризуется показателями (ТРБД-02-01) – (ТРБД-02-03), в настоящее время в явном виде представлено единственным стандартом [25], вступившим в действие с 1 ноября 2021 г., да и то с отставанием в несколько лет по сравнению с идентичным международным стандартом. Отдельные моменты работы с массивами данных, используемых для обучения, верификации и валидации в отдельных сферах деятельности, затрагиваются несколькими стандартами в области искусственного интеллекта (см. перечень на веб-сайте технического комитета по стандартизации «Искусственный интеллект» [33]).

Можно констатировать, что нормативное техническое регулирование работы с большими данными в Российской Федерации находится на начальном уровне развития, и оценка перечисленных выше показателей – *неудовлетворительно*.

Походы к саморегулированию при работе с большими данными (показатель СРБД-02-01) в Российской Федерации были инициированы в 2019 г. Ассоциацией больших данных, разработавшей совместно с Институтом развития интернета Кодекс этики использования данных [34], и призывающей других участников рынка больших данных присоединиться к нему. Принятие данного документа и присоединение к нему – попытка восполнить недостатки нормативного регулирования в стране. Кодекс не является обязательным документом даже для присоединившихся к нему организаций и не устанавливает ответственности за его несоблюдение, являясь, по сути дела, декларацией о намерениях, что несколько снижает его значимость. Итоговая оценка уровня саморегулирования в Российской Федерации – *хорошо*.

Нормативное регулирование процессов обеспечения доверия и информационной безопасности при работе с большими данными (показатель ДББД-02-01) представлено достаточно широко как на законодательном уровне, так и на уровне подзаконных актов, в первую очередь благодаря актам, направленным на безопасное использование персональных данных. Что же касается нормативного регулирования актуальности, полноты, доверия к большим данным, то подобные новеллы в российском законодательстве отсутствуют. В лучшем случае такие вопросы фиксируются в договорных отношениях и подпадают под гражданско-правовое регулирование. При этом сохраняется высокий риск представления поддельных и/или неактуальных данных в качестве доверенных данных.

Подводя итог, можно оценить ситуацию с нормативным регулированием обеспечения доверия и информационной безопасности при работе с большими данными Российской Федерации как *удовлетворительную*.

Заключение

Сегодня крайне важно сформировать государственную политику Российской Федерации по работе с данными, выработать инструменты для реализации такой политики.

Задача формирования национальной политики и ее реализация – это та первоочередная задача, которую сейчас нужно решать, в том числе в рамках разрабатываемой сегодня стратегии социально-экономического развития России до 2030 года. По целому ряду направлений в явном виде нужно учитывать возможности, которые несут за собой цифровые технологии и данные, обрабатываемые с их помощью. При этом вопрос, связанный с национальной безопасностью и управлением промышленными данными, чем дальше, тем больше будет выходить на первый план как в Российской Федерации, так и во всем мире. Поэтому наличие целенаправленной государственной политики, подкрепленной соответствующей национальной стратегией и планом действий по ее реализации, – один из вызовов, которые стоят перед Россией.

В настоящей работе описана методология мониторинга и оценки государственной политики и регулирования работы с большими данными в России.

Применимость предлагаемой концептуальной схемы и вошедших в нее показателей мониторинга доказана в рамках пилотной реализации с учетом рекомендованных доработок.

Дальнейшее развитие настоящего исследования может идти в направлении уточнения концептуальной схемы мониторинга, а также уточнения и пересмотра перечня показателей, который может потребоваться после сбора всех необходимых данных для проведения оценки, а также получения обратной связи от опрашиваемых организаций и профессионального сообщества.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
2. Всемирный банк. 2018 год. Доклад о развитии цифровой экономики в России, сентябрь 2018 года, «Конкуренция в цифровую эпоху: стратегические вызовы для Российской Федерации».
3. Ershova T.V., Hohlov Yu.E., Shaposhnik S.B. "Methodology for Digital Economy Development Assessment as a Tool for Managing the Digital Transformation Processes," 2018 Eleventh International Conference "Management of large-scale system development" (MLSD, Moscow, Russia, 2018). – P. 1-3. <https://doi.org/10.1109/MLSD.2018.8551846>
4. Указ Президента Российской Федерации «О национальных целях развития Российской Федерации на период до 2030 года» от 21 июля 2020 года № 474.
5. Ершова Т.В., Хохлов Ю.Е., Шапошник С.Б. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
6. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government, *OECD Public Governance Policy Papers*, No. 02, OECD Publishing, Paris, <https://doi.org/10.1787/f64fed2a-en>
7. Alley-Young G. (2017) White House Big Data Initiative. In: Schintler L., McNeely C. (eds) Encyclopedia of Big Data. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_204-1
8. Big data: Seizing opportunities, Preserving values. May 2014. URL: https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf (дата обращения 01.06.2021)
9. The Federal Big Data Research and Development Strategic Plan. URL: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/bigdatardstrategicplan-nitrd_final-051916.pdf (дата обращения 01.06.2021)
10. Federal Data Strategy. URL: <https://strategy.data.gov/> (дата обращения 01.06.2021)
11. Federal Data Strategy. 2020 Action Plan. URL: <https://strategy.data.gov/action-plan/> (дата обращения 01.06.2021)
12. NIST Big Data interoperability Framework. V1.0 Final Version. URL: https://bigdatawg.nist.gov/V1_output_docs.php (дата обращения 01.06.2021)
13. NIST Big Data interoperability Framework. V2.0 Final Version. URL: https://bigdatawg.nist.gov/V2_output_docs.php (дата обращения 01.06.2021)
14. NIST Big Data interoperability Framework. V3.0 Final Version. URL: https://bigdatawg.nist.gov/V3_output_docs.php (дата обращения 01.06.2021)
15. Аверкин А.Н., Афанасьев С.Д., Микрюков А.А., Паджев В.В., Райков А.Н., Хохлов Ю.Е., Храмовская Н.А. Стандартизация работы с большими данными: международные и национальные стандарты // Информационное общество. 2021. № 4–5. С. 220–258. https://doi.org/10.52605/16059921_2021_04_220
16. European Commission. A European Strategy for Data, Brussels, 19.2.2020, COM(2020) 66 final (Feb. 19, 2020). URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN> (дата обращения 01.06.2021)
17. European Commission. General Data Protection Regulation. URL: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> (дата обращения 01.06.2021)
18. Data Strategy of the German Federal Government. URL: <https://www.bundesregierung.de/breg-en/service/information-material-issued-by-the-federal-government/data-strategy-of-the-federal-german-government-1950612> (дата обращения 01.06.2021)
19. UK data capability strategy: seizing the data opportunity. URL: <https://www.gov.uk/government/publications/uk-data-capability-strategy> (дата обращения 01.06.2021)
20. UK National Data Strategy. URL: <https://www.gov.uk/government/publications/uk-national-data-strategy> (дата обращения 01.06.2021)

21. Global Britain in a Competitive Age: The Integrated Review of Security, Defence, Development and Foreign Policy. URL: <https://www.gov.uk/government/publications/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy> (дата обращения 01.06.2021)
22. China academy of information and communications technology. Big Data White Paper (2019). URL: http://www.caict.ac.cn/english/research/whitepapers/202003/t20200327_278200.html (дата обращения 01.06.2021)
23. Data Security Law of the PRC. URL: <https://www.chinalawtranslate.com/en/datasecuritylaw/> (дата обращения 01.06.2021)
24. Аналитический центр при Правительстве Российской Федерации. Основные результаты анализа законодательства зарубежных государств и Российской Федерации в области управления данными, описания моделей правового регулирования управления данными. URL: <https://ac.gov.ru/news/page/eksperty-izucili-zarubeznyj-opyt-v-regulirovanii-upravlenia-gosdannymi-26589> (дата обращения 01.06.2021)
25. ГОСТ Р ИСО/МЭК 20546–2021 «Информационные технологии. Большие данные. Обзор и словарь». URL: <http://protect.gost.ru/document1.aspx?control=31&baseC=6&page=4&month=4&year=-1&search=&id=240981> (дата обращения 01.06.2021)
26. Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы. Утверждена Указом Президента Российской Федерации от 9 мая 2017 г. № 203.
27. Перечень инициатив социально-экономического развития Российской Федерации до 2030 года. Утвержден распоряжением Правительства Российской Федерации от 06.10.2021 № 2816-р.
28. Дмитрий Чернышенко: Президиум правкомиссии одобрил стратегии цифровой трансформации российской экономики. URL: <http://government.ru/news/42623/> (дата обращения 01.10.2021)
29. Стратегическое направление в области цифровой трансформации государственного управления. Утверждено распоряжением Правительства Российской Федерации от 22 октября 2021 г. № 2998-р.
30. Государственная программа Российской Федерации «Информационное общество» URL: <https://digital.gov.ru/ru/activity/programs/1/> (дата обращения 01.10.2021)
31. Цифровая экономика Российской Федерации. URL: <https://digital.gov.ru/ru/activity/directions/858/> (дата обращения 01.10.2021)
32. ФГИС Ки. Ведомственные программы цифровой трансформации. URL: https://portal.eskigov.ru/documents?category=vpct_programs (дата обращения 01.10.2021)
33. Технический комитет по стандартизации «Искусственный интеллект» (ТК 164). URL: <https://www.tc164.ru/утвержденные-стандарты> (дата обращения 01.10.2021)
34. Кодекс этики использования данных. Ассоциация больших данных, Институт развития интернета, 2019. URL: <https://rubda.ru/deyatelnost/kodeks/> (дата обращения 01.06.2021)

BIG DATA GOVERNMENT POLICY AND REGULATION

Orlov, Stepan Vladimirovich

Candidate of economic sciences, associate professor

Lomonosov Moscow State University, head of Department of history of social movements and political parties

Moscow, Russian Federation

faculty@hist.msu.ru

Padzhev, Valentin Valentinovich

Institute of the Information Society, head of Directorate of legal programs

Moscow Russian Federation

vpadzhev@iis.ru

Hohlov, Yuri Evgenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor

Moscow Russian Federation

yuri.hohlov@iis.ru

Abstract

A framework and a set of indicators developed for monitoring public policy and regulation in the field of big data as a factor influencing the development, production, and use of big data technologies are presented. The framework includes indicators characterizing strategic planning, legal and technical regulation, self-regulation and regulation of processes to ensure trust and security in working with big data. Methods for measuring indicators for monitoring public policy and regulating work with big data are proposed, and the results of measuring available indicators are presented.

Keywords

big data technologies; government policy for big data; big data regulation; big data technical regulation; big data self-regulation; trust and security in working with big data; Big Data for Digital Economy; BD4DE

References

1. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
2. Vsemirnyj bank. 2018 god. Doklad o razvitiij cifrovoj ekonomiki v Rossii, sentyabr' 2018 goda, «Konkurenciya v cifrovuyu epohu: strategicheskie vyzovy dlya Rossijskoj Federacii».
3. Ershova T.V., Hohlov Yu.E., Shaposhnik S.B. Methodology for Digital Economy Development Assessment as a Tool for Managing the Digital Transformation Processes/ 2018 Eleventh International Conference "Management of large-scale system development" (MLSD, Moscow, Russia, 2018). – P. 1-3. <https://doi.org/10.1109/MLSD.2018.8551846>
4. Ukaz Prezidenta Rossijskoj Federacii «O nacional'nyh celyah razvitiya Rossijskoj Federacii na period do 2030 goda» ot 21 iyulya 2020 goda № 474.
5. Ershova T.V., Hohlov Yu.E., Shaposhnik S.B. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologij raboty s bol'shimi dannymi // Informacionnoe obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
6. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government, OECD Public Governance Policy Papers, No. 02, OECD Publishing, Paris, <https://doi.org/10.1787/f64fed2a-en>
7. Alley-Young G. (2017) White House Big Data Initiative. In: Schintler L., McNeely C. (eds) Encyclopedia of Big Data. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_204-1
8. Big data: Seizing opportunities, Preserving values. May 2014. URL: https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf (accessed 01.06.2021)

9. The Federal Big Data Research and Development Strategic Plan. URL: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/bigdatardstategicplan-nitrtd_final-051916.pdf (accessed 01.06.2021)
10. Federal Data Strategy. URL: <https://strategy.data.gov/> (accessed 01.06.2021)
11. Federal Data Strategy. 2020 Action Plan. URL: <https://strategy.data.gov/action-plan/> (accessed 01.06.2021)
12. NIST Big Data interoperability Framework. V1.0 Final Version. URL: https://bigdatawg.nist.gov/V1_output_docs.php (accessed 01.06.2021)
13. NIST Big Data interoperability Framework. V2.0 Final Version. URL: https://bigdatawg.nist.gov/V2_output_docs.php (accessed 01.06.2021)
14. NIST Big Data interoperability Framework. V3.0 Final Version. URL: https://bigdatawg.nist.gov/V3_output_docs.php (accessed 01.06.2021)
15. Averkin A.N., Afanas'ev S.D., Mikryukov A.A., Padzhev V.V., Rajkov A.N., Hohlov Yu.E., Hramcovskaya N.A. Standartizaciya raboty s bol'shimi dannymi: mezhdunarodnye i nacional'nye standarty // Informacionnoe obshchestvo. 2021. № 4–5. S. 220–258. URL: https://doi.org/10.52605/16059921_2021_04_220
16. European Commission. A European Strategy for Data, Brussels, 19.2.2020, COM(2020) 66 final (Feb. 19, 2020). URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN> (accessed 01.06.2021)
17. European Commission. General Data Protection Regulation. URL: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> (accessed 01.06.2021)
18. Data Strategy of the German Federal Government. URL: <https://www.bundesregierung.de/breg-en/service/information-material-issued-by-the-federal-government/data-strategy-of-the-federal-german-government-1950612> (accessed 01.06.2021)
19. UK data capability strategy: seizing the data opportunity. URL: <https://www.gov.uk/government/publications/uk-data-capability-strategy> (accessed 01.06.2021)
20. UK National Data Strategy. URL: <https://www.gov.uk/government/publications/uk-national-data-strategy> (accessed 01.06.2021)
21. Global Britain in a Competitive Age: The Integrated Review of Security, Defence, Development and Foreign Policy. URL: <https://www.gov.uk/government/publications/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy> (accessed 01.06.2021)
22. China academy of information and communications technology. Big Data White Paper (2019). URL: http://www.caict.ac.cn/english/research/whitepapers/202003/t20200327_278200.html (accessed 01.06.2021)
23. Data Security Law of the PRC. URL: <https://www.chinalawtranslate.com/en/datasecuritylaw/> (accessed 01.06.2021)
24. Analiticheskij centr pri Pravitel'stve Rossijskoj Federacii. Osnovnye rezul'taty analiza zakonodatel'stva zarubezhnyh gosudarstv i Rossijskoj Federacii v oblasti upravleniya dannymi, opisaniya modelej pravovogo regulirovaniya upravleniya dannymi. URL: <https://ac.gov.ru/news/page/eksperty-izucili-zarubezhnyj-opyt-v-regulirovanii-upravlenia-gosdannymi-26589> (accessed 01.06.2021)
25. GOST R ISO/MEK 20546–2021 «Informacionnye tekhnologii. Bol'shie dannye. Obzor i slovar'». URL: <http://protect.gost.ru/document1.aspx?control=31&baseC=6&page=4&month=4&year=-1&search=&id=240981> (accessed 01.06.2021)
26. Strategiya razvitiya informacionnogo obshchestva v Rossijskoj Federacii na 2017–2030 gody. Utverzhdena Ukazom Prezidenta Rossijskoj Federacii ot 9 maya 2017 g. № 203.
27. Perechen' iniciativ social'no-ekonomicheskogo razvitiya Rossijskoj Federacii do 2030 goda. Utverzhden rasporyazheniem Pravitel'stva Rossijskoj Federacii ot 06.10.2021 № 2816-r.
28. Dmitrij Chernyshenko: Prezidium pravkomissii odobril strategii cifrovoj transformacii rossijskoj ekonomiki. URL: <http://government.ru/news/42623/> (accessed 01.10.2021)
29. Strategicheskoe napravlenie v oblasti cifrovoj transformacii gosudarstvennogo upravleniya. Utverzhdeno rasporyazheniem Pravitel'stva Rossijskoj Federacii ot 22 oktyabrya 2021 g. № 2998-r.
30. Gosudarstvennaya programma Rossijskoj Federacii «Informacionnoe obshchestvo» URL: <https://digital.gov.ru/ru/activity/programs/1/> (accessed 01.10.2021)

31. Cifrovaya ekonomika Rossijskoj Federacii. URL:
<https://digital.gov.ru/ru/activity/directions/858/> (accessed 01.10.2021)
32. FGIS KI. Vedomstvennye programmy cifrovoj transformacii. URL:
https://portal.eskigov.ru/documents?category=vpct_programs (accessed 01.10.2021)
33. Tekhnicheskij komitet po standartizacii «Iskusstvennyj intellekt» (TK 164). URL:
<https://www.tc164.ru/utverzhdennye-standarty> (accessed 01.10.2021)
34. Kodeks etiki ispol'zovaniya dannyh. Associaciya bol'shih dannyh, Institut razvitiya interneta, 2019. URL: <https://rubda.ru/deyatelnost/kodeks/> (accessed 01.06.2021)

Информационное общество: политика и факторы развития

СИСТЕМА УПРАВЛЕНИЯ И МЕХАНИЗМЫ ФИНАНСИРОВАНИЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Елизаров Александр Михайлович

*Доктор физико-математических наук, профессор
Институт информационных технологий и интеллектуальных систем Казанского (Приволжского)
федерального университета, Кафедра программной инженерии, профессор
Казань, Российская Федерация
amelizarov@gmail.com*

Паджев Валентин Валентинович

*Институт развития информационного общества, руководитель Дирекции правовых программ
Москва, Российская Федерация
vpadzhev@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель Совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель Базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Представлены концептуальная схема и набор показателей, разработанные для мониторинга системы управления и механизмов финансирования работы с большими данными как одного из факторов, влияющих на, производство, использование и развитие технологий работы с большими данными. В концептуальную схему включены два компонента, охватывающих саму систему управления развитием и использованием технологий работы с большими данными, а также механизмы финансирования этих работ. Основными показателями определены наличие и эффективность системы управления работой с большими данными; участие заинтересованных сторон в этих процессах; наличие и состояние системы мониторинга работы с большими данными и ее соответствие международным системам; цифровое лидерство в области работы с большими данными и международное сотрудничество в этой области; источники, объемы и модели финансирования работы с большими данными; стимулирование и обеспеченность финансированием деятельности по работе с большими данными.

Ключевые слова

большие данные; технологии работы с большими данными; система управления работой с большими данными; механизмы финансирования работы с большими данными; эффективность системы управления работой с большими данными; участие заинтересованных сторон; цифровое лидерство; модели финансирования работы с большими данными; Big Data for Digital Economy; BD4DE

Введение

Использование цифровых технологий работы с большими данными дает возможность получения принципиально новых результатов (по сравнению с применением традиционных ИКТ), открывает новые ниши, а иногда и новые рынки, создает мощнейшую конкуренцию среди разработчиков и поставщиков услуг.

© Елизаров А.М., Паджев В.В., Хохлов Ю.Е., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_53

Обратной стороной массовой увлеченности попытками занять в экономике данных лучшие места является расходование большого количества ресурсов, существенный объем которых часто тратится безрезультатно. Поэтому возникает необходимость концентрации ресурсов на наиболее важных направлениях развития и применении технологий, обладающих наибольшей эффективностью использования. При этом, естественно, должна обеспечиваться конкуренция как участников рынка, так и технологий. На национальном (или отраслевом) уровне это требует наличия системы управления, которая может обеспечить целенаправленное воздействие на процессы развития и использования технологий. Слабое управление приводит к плохой координации, отсутствию коммуникаций, нарушению сроков и в итоге к плохим результатам. С другой стороны, эффективное управление позволяет организовать разработку и реализацию политики и стратегии научно-технологического развития, в том числе на основе технологий работы с большими данными.

В докладе Всемирного банка «Разрубая гордиев узел данных» [1] надежная и обеспеченная ресурсами институциональная среда обозначена как один из ключевых факторов, который имеет решающее значение для развития экономики данных, поддерживаемой цифровыми технологиями. В нее включаются как «стимулирующие» институты, которые отвечают за разработку и реализацию стратегий, политики, законов, нормативных актов, стандартов и руководящих принципов, так и «защитные» институты, которые контролируют и мониторят прогресс, обеспечивают соблюдение правил, а также предлагают гражданам соразмерное и эффективное возмещение. В этом случае система управления должна обеспечивать развитие и использование технологий работы с большими данными в интересах социально-экономического развития страны, в том числе путем вовлечения в процессы управления представителей основных заинтересованных сторон (власти, бизнеса, гражданского общества, научно-образовательного сообщества).

В Российской Федерации управление процессами развития цифровой экономики еще больше усложняется в связи с необходимостью привлечения к управлению не только перечисленных выше основных заинтересованных сторон, но и представителей регионов и муниципалитетов. Структуры управления в стране (или отдельной сфере деятельности) должны быть прозрачными, инклюзивными, федеративными по своей сути и находиться под централизованным управлением.

Система управления работой с (большими) данными должна быть осознана высшим руководством как в государственном, так и в частном секторах. Определяющей характеристикой на национальном уровне является устойчивая политическая воля высшего руководства при решении проблем цифровизации и информирования о важности данных как инструмента обслуживания граждан своей страны, что подразумевает фундаментальные культурные изменения, основанные на философии совместного и повторного использования данных. Крайне важно, чтобы такой подход был одобрен и поддержан высшим руководством и такая поддержка сохранялась на протяжении всего периода ее реализации.

Эффективная система управления должна обеспечивать:

- согласование интересов заинтересованных сторон;
- прозрачность и ясность процесса принятия решений;
- мониторинг и оценку осуществляемых действий и ожидаемых результатов;
- коммуникации между заинтересованными сторонами по вопросам обмена лучшими практиками и инсайтами, извлечение уроков для достижения наибольшего эффекта от использования данных;
- соблюдение нормативных правовых требований, касающихся всех этапов жизненного цикла работы с данными;
- следование стандартам, корпоративным моделям данных, общим методологиям разработки, концептуальным схемам интероперабельности;
- сотрудничество и сотворчество в предоставлении инновативных совместных услуг;
- приоритет защиты данных и кибербезопасности

и многое другое.

Система управления сама по себе не может обеспечить или гарантировать что-либо, а нужный результат может быть получен лишь путем применения инструментов, которые реализуют усилия и решают задачи, формируемые системой управления, в виде конкретных действий. Одними из наиболее важных инструментов управления являются механизмы финансирования,

целенаправленность и гибкость которых могут обеспечить решение большинства задач управления. На выбор механизмов финансирования оказывают существенное влияние ограниченность ресурсов, корректность поставленной задачи и сопровождающие ее риски, ожидание отдачи (результата) и многое другое, что выходит за рамки настоящего исследования.

Как отмечено выше, наличие зрелой системы управления процессами работы с большими данными формирует благоприятные условия для цифровой трансформации экономики и общества, поэтому регулярный мониторинг институциональной среды, проводимый на национальном, региональном или отраслевом уровнях, позволит оценить влияние этого фактора и сформировать рекомендации по доработке действующей системы управления и финансовых механизмов воздействия на процессы развития экономики данных.

В статье представлены компоненты концептуальной схемы мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации BD4DE (Big Data for Digital Economy) [2], относящиеся к системе управления и механизмам финансирования работы с большими данными.

1 Анализ подходов к оценке системы управления и механизмов финансирования работы с большими данными

Большинство стран-лидеров развития цифровой экономики отвечает важность совершенствования системы управления и механизмов финансирования работы с большими данными.

В Докладе о мировом развитии 2021 «Данные для лучшей жизни» [3] подчеркнуты важность формирования экосистемы данных, включающей в себя поставщиков данных и услуг, разработчиков, потребителей данных, а также институциональных структур для успешного перехода к экономике, управляемой на основе данных. При этом обращено внимание на необходимость при разработке соответствующих структур управления одновременного принятия во внимание местных условий.

Опыт стран-лидеров развития цифровой экономики показывает, что в центре эффективной системы управления работой с данными (в том числе большими), которая гарантированно обеспечивает реализацию выбранной стратегии и достижение намеченных результатов, находятся структуры, роли, принципы, политика и практика. При этом структуры управления должны быть адаптированы к конкретным потребностям и уровню зрелости органа власти или крупной организации частного сектора. При этом большое значение всегда имеет вовлеченность первых лиц: система управления работой с данными должна быть одобрена и поддержана высшим руководством, а также находиться под их постоянным контролем. Одной из форм проявления контроля служит назначение на самом высоком уровне государств и организаций руководителей, ответственных за реализацию стратегии и вообще за работу с данными

В частности, эффективная система управления работой с данными во многих национальных стратегиях работы с данными, как правило, включает в себя в качестве основных компонентов:

- организационные структуры, в которые входят представители всех заинтересованных сторон: органов власти, бизнеса; гражданского общества, научно-образовательного сообщества;
- четко определенные роли и обязанности;
- прозрачные процедуры принятия решений;
- процедуры национальной системы менеджмента данных, особенно в отношении общих пространств данных, необходимых для развития экономики данных и используемых как органами власти, так и частными компаниями различного масштаба;
- соглашения об уровне обслуживания между поставщиками данных, услуг и инфраструктуры;
- международные соглашения по обеспечению интероперабельности, облегчающие трансграничную передачу данных.

Участие представителей основных заинтересованных сторон означает, что федеративная модель управления работой с данными является наиболее подходящей для реализации национальной стратегии в области работы с данными с централизованным органом управления и децентрализованными организациями государственного и частного секторов на национальном, отраслевом, региональном и муниципальном уровнях.

Организация экономического сотрудничества и развития (ОЭСР) констатирует [4], что надежная система управления является ключевым фактором, характеризующим зрелое цифровое правительство, и рекомендует всем странам укрепить институциональный потенциал для обеспечения реализации национальной цифровой стратегии и стратегии работы с данными.

Упомянутый ранее доклад Всемирного банка [1] определил надежные и обеспеченные ресурсами институты как один из пяти компонентов, необходимых правительствам, стремящимся обеспечить надежное совместное использование данных, что имеет решающее значение для экономики данных. Анализ ситуации с цифровым развитием в семи странах (Индия, Эстония, Сингапур, Чили, Маврикий, Уругвай и Мексика), представленный в этом докладе, показал, что модернизация институциональной среды является одним из факторов успеха. Аналогичным образом некоторые другие страны адаптировали полномочия существующих государственных органов или создали новые органы для реализации стратегии в области работы с данными и создания соответствующих структур управления. Со временем, когда правительства осознали важность цифровой трансформации, политический вес этих структур значительно возрос, и они переместились с периферии в центр правительства.

Во Франции и Соединенном Королевстве ответственность за цифровые вопросы лежит на премьер-министрах, в то время как в федеральном правительстве Германии есть государственный министр по цифровизации, который одновременно отвечает за координацию деятельности регионов страны по цифровым вопросам. Соединенное Королевство планирует назначить отдельного руководителя по работе с данными для контроля реализации своей новой национальной стратегии работы с данными: он будет работать при поддержке межведомственных органов управления, включая орган по стандартизации работы с данными.

В Соединенных Штатах Америки согласно плану действий [5] по реализации принятой недавно Федеральной стратегии работы с данными [6] на 2020 год федеральным руководителям по работе с данными поручено возглавить процесс внедрения изменений в своих организациях и повысить качество данных.

Наконец, несколько стран прямо увязали цифровую трансформацию с экономикой или инновациями. Например, в Италии это Министерство инновационных технологий и цифровизации, в Испании же деятельность Министерства экономики и цифровой трансформации курирует вице-премьер.

Таким образом, подавляющим большинством развитых стран признана важность совершенствования системы управления работой с большими данными.

Остановимся теперь на ситуации с организацией управления работой с данными в Российской Федерации. В отличие от других национальных проектов реализация национальной программы «Цифровая экономика Российской Федерации» осуществляется при обязательном участии и координации действий между «ключевыми игроками» цифровой экосистемы страны, в которую входят органы власти, бизнес и научно-образовательное сообщество [7]. Федеральным органом исполнительной власти, осуществляющим координацию действий ведомств, участвующих в реализации этой национальной программы, является Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. Координацию действий других заинтересованных сторон, в первую очередь бизнеса и научно-образовательного сообщества, осуществляет некоммерческая организация «Цифровая экономика», учрежденная в 2017 году специально для этих целей [8]. Хотя такая модель управления является достаточно инклюзивной и эффективной, в том числе для управления работой с данными, в ней не представлены пользователи и гражданское общество, что снижает потенциал учета интересов основных заинтересованных сторон. Описанная модель также не действует на отраслевом, региональном или муниципальном уровнях управления, что затрудняет качественное управление работой с данными в регионах, а также взаимодействие между регионами и федеральным центром.

Таким образом, адаптация в России существующих лучших практик управления данными, рассмотренных выше, позволила бы ей существенно продвинуться вперед в процессе перехода к экономике данных, включая цифровую трансформацию государственного сектора, который должен стать «цифровым по умолчанию», а также основанным на работе с данными и ориентированным на пользователей. Для полного представления об имеющейся ситуации и определения перспективных направлений развития системы управления и механизмов

финансирования работы с большими данными необходим постоянный мониторинг названной предметной области.

В Российской Федерации для финансирования проектов, в том числе в области больших данных, используют различные механизмы финансирования, в частности, бюджетное, частное и финансирование в рамках различных форм государственно-частных партнерств. Применяют также различные формы финансовой поддержки, например, в виде субсидий, грантов, налоговых льгот (налоговых маневров, сниженных (льготных) ставок налогообложения, льгот налогообложения для резидентов особых экономических зон, технопарков), льготных кредитов, в том числе путем предоставления компенсаций кредитных ставок. Активно развиваются венчурное финансирование и финансирование институтами развития. Однако все механизмы (источники, модели, условия) финансирования в разной степени ограничивают реализацию проектов. Основным ограничением, конечно, являются отсутствие или существенные трудности доступа к источникам финансирования, которые определяются как особенностями (условиями) финансирования самих проектов, так и отсутствием ресурсов, соответствующих имеющимся в них потребностям. Это приводит к необходимости формирования сложных моделей и механизмов, обеспечивающих вовлечение в финансирование нескольких источников.

Стремление государства облегчить доступность финансирования и стимулировать, в частности, развитие технологий работы с большими данными при ограниченных объемах финансирования реализуется в множестве форм поддержки. В итоге выбор механизма финансирования определяется целым комплексом факторов, к которым, например, относятся:

- организационно-правовая форма потребителя финансирования;
- степень участия государства в финансировании;
- реализуемость проекта;
- состав потребителей результатов реализованного проекта;
- эффекты от реализации проекта;
- требуемые объемы финансирования проекта;
- наличие источника финансирования, соответствующего проекту;
- риски ограничения конкуренции в результате выбора неправильных источника финансирования и соответствующих ему условий;
- риски нереализации проекта

и ряд других.

Одновременно последовательное решение задачи цифровизации экономики предполагает четкое целеполагание и увязку целевой установки на цифровизацию экономики с основными мероприятиями по ее достижению, закрепленными в соответствующих документах стратегического планирования, таких, например, как национальный проект «Цифровая экономика», базовые стратегии социально-экономического и научно-технологического развития, отраслевые стратегии развития, а также государственные программы в сфере цифровой экономики и другие национальные проекты. Все эти программы, естественно, должны содержать планы и некие механизмы финансирования названных мероприятий.

Отметим, в что стратегии развития отечественного рынка больших данных до 2024 года, представленной Ассоциацией больших данных [9], кроме определения границы рынка больших данных, указаны 5 сценариев развития технологий больших данных (от пессимистичного до сценария мечты), среди которых обязательными для реализации в России названы и описаны базовый и оптимистичный сценарии, причем только в последнем в явном виде обозначены вопросы финансирования.

2 Концептуальная схема мониторинга и система показателей

Результаты анализа, представленные в предыдущем разделе, позволили описать концептуальную схему мониторинга системы управления и механизмов финансирования работы с большими данными в виде двух компонентов, охватывающих саму систему управления развитием и использованием технологий работы с большими данными, а также механизмы финансирования развития и использования технологий работы с большими данными (см. рисунок 1).

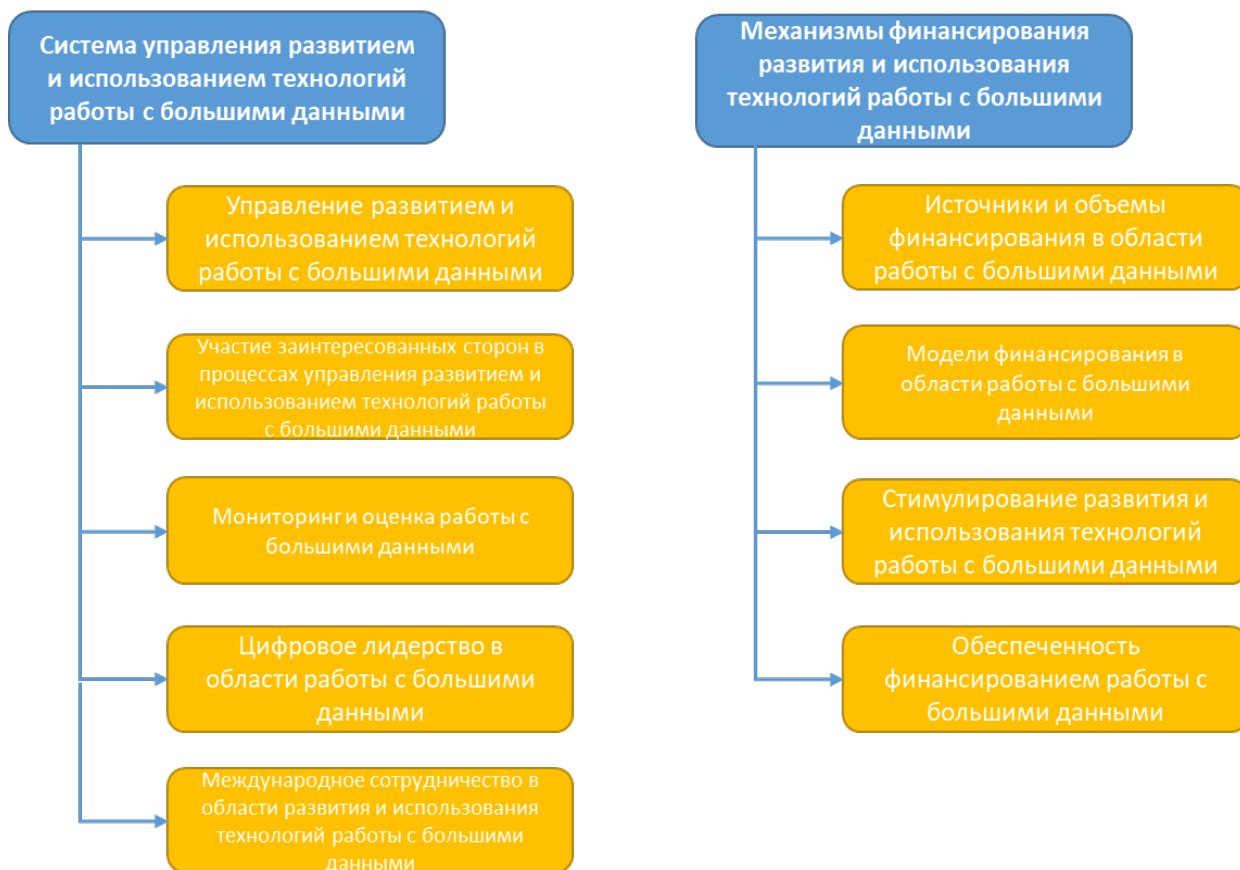


Рисунок 1. Концептуальная схема предметной области мониторинга «Система управления и механизмы финансирования работы с большими данными»

Система показателей для мониторинга рассматриваемой предметной области в целом основана на показателях, ранее апробированных при оценке текущего состояния развития цифровой экономики в России (см. [10]). Ниже приведен перечень выбранных показателей, характеризующих каждую из двух выделенных предметных областей мониторинга, а также охарактеризовано их смысловое содержание.

2.1 Система управления развитием и использованием технологий работы с большими данными

Мониторинг системы управления развитием и использованием технологий работы с большими данными на национальном (или отраслевом) уровне основан на оценке нескольких основных аспектов, таких как наличие и эффективность функционирования всей системы управления, участие в процессах управления основных заинтересованных сторон; наличие системы мониторинга в составе системы управления, цифровое лидерство и международное сотрудничество в данной сфере.

Для характеристики процессов управления развитием и использованием технологий работы с большими данными целесообразно оценить и сопоставить между собой два показателя:

- (СУРБД-01-01) Наличие и эффективность системы управления развитием цифровой экономики
- (СУРБД-01-02) Наличие и эффективность системы управления развитием и использованием технологий работы с большими данными

При этом управление развитием и использование технологий работы с большими данными выступают в качестве самостоятельных объектов управления. Применение двух показателей позволяет охарактеризовать наличие и эффективность системы управления развитием и использованием различных цифровых технологий для решений задач социально-экономического развития и сопоставить с ней систему управления работой с большими данными.

Еще одной характеристикой эффективной системы управления является

- (СУРБД-01-03) *Участие заинтересованных сторон в процессах управления развитием и использованием технологий работы с большими данными*

При расчетах этого показателя нужно учитывать наличие представителей основных заинтересованных сторон в системе управления, а также стадии процесса управления, в которых принимают участие представители названных сторон. Прежде всего, необходимо оценить, насколько вовлечены в процессы управления развитием и использованием технологий работы с большими данными представители органов власти, бизнеса, гражданского общества и научно-образовательного сообщества. Далее, если их участие будет подтверждено, нужно оценить, насколько полно они представлены на различных стадиях развития и использования технологий работы с большими данными, таких как разработка нормативного регулирования; принятие решений о мерах поддержки; мониторинг и оценка достигнутых результатов.

Существенным элементом любой системы управления является наличие «петли обратной связи» для принятия управляющих воздействий, реализуемой благодаря регулярному мониторингу и оценке происходящих процессов. В данном случае этим целям служат два показателя:

- (СУРБД-01-04) *Наличие и состояние системы мониторинга работы с большими данными*
- (СУРБД-01-05) *Соответствие системы мониторинга работы с большими данными международным системам*

При расчете первого из них важно получить ответы на основной вопрос – насколько полно на каждом из этапов работы с большими данными ведутся мониторинг и оценка:

- прогноза развития и использования технологий работы с большими данными;
- потребностей в использовании технологий работы с большими данными;
- уровня разработки технологий работы с большими данными;
- использования технологий работы с большими данными;
- достигнутых эффектов от использования технологий работы с большими данными.

При расчете второго показателя необходимо получить ответ на вопрос, в какой степени мониторинг и оценка работы с большими данными на национальном уровне соответствуют лучшим международным практикам.

Анализ существующих подходов к построению систем управления процессами технологического развития показал необходимость учета одного из ключевых факторов влияния на анализируемые процессы:

- (СУРБД-01-06) *Цифровое лидерство в области работы с большими данными,*

который характеризует наличие в системе управления и персональную ответственность руководителей высокого уровня за развитие и использование технологий работы с большими данными.

Наконец, еще одной важной характеристикой эффективной системы управления является участие в международной кооперации, открывающее возможности использования лучших международных практик, накопленных в других странах. Для этого будем использовать показатель

- (СУРБД-01-07) *Международное сотрудничество в области развития и использования технологий работы с большими данными*

При расчете этого показателя будут учитываться не только наличие сотрудничества (программ, проектов, инициатив) по обмену знаниями с другими странами в области развития и использования технологий работы с большими данными, но и соответствие международного сотрудничества целям и задачам развития и использования отечественных технологий работы с большими данными. Необходимо учитывать также и направления международного сотрудничества основных заинтересованных сторон, прежде всего, власти, научно-образовательного сообщества и бизнеса.

2.2 Механизмы финансирования развития и использования технологий работы с большими данными

Мониторинг механизмов финансирования развития и использования технологий работы с большими данными основан на оценке следующих отдельных компонентов: источники и объемы финансирования в области работы с большими данными; модели финансирования в области работы с большими данными; стимулирование развития и использования технологий работы с

большими данными; обеспеченность работы с большими данными соответствующим финансированием.

При оценке показателя

- (МФБД-02-01) *Источники и объемы финансирования в области работы с большими данными*

предлагается определить, в какой степени для развития и использования технологий работы с данными используется финансирование из перечисленных ниже источников:

- бюджетная система государства;
- бюджеты частного сектора;
- бюджеты физических лиц;
- бюджеты институтов развития.

Такая оценка позволит заключить, в какой степени (в диапазоне «совсем не обеспечено – обеспечено в необходимом объеме») обеспечены финансированием разработка технологий и решений для работы с большими данными, а также их использование в различных сферах деятельности.

Для оценки основных подходов к финансированию работы с большими данными будет использован показатель

- (МФБД-02-02) *Модели финансирования в области работы с большими данными,*

который будет оцениваться в двух разрезах: по степени рисков (агрессивная, компромиссная, консервативная модели) и по отношению к результатам (финансирование по результату, «для поддержки», «от достигнутого», проектное, выдающихся проектов, прочее прямое и целевое).

Для оценки мер стимулирования работы с большими данными используется показатель

- (МФБД-02-03) *Стимулирование развития и использования технологий работы с большими данными,*

который будет оцениваться по таким аспектам, как стимулирование конкуренции, разработки и использования отечественных технологий работы с большими данными, инноваций, невостребованных технологий работы с большими данными.

Наконец, последний показатель по данному направлению мониторинга

- (МФБД-02-04) *Обеспеченность финансированием деятельности по работе с большими данными*

будет характеризовать наличие средств для деятельности как по развитию, так и по использованию технологий работы с большими данными в различных сферах деятельности.

3 Методология исследования

В набор методологических инструментов для мониторинга факторов, влияющих на производство, использование и воздействие технологий работы с большими данными в различных аспектах, в соответствии с [2] входят кабинетные исследования (оценка документов, сбор и обработка данных из различных источников), опросы организаций и экспертные опросы. Для рассматриваемых нами системы управления и механизмов финансирования развития и использования технологий работы с большими данными основными методами исследования стали кабинетные и экспертные исследования и оценки. Для их реализации необходимо было разработать специализированные анкеты для опроса, в которых нужно было использовать показатели, прошедшие экспертную оценку. В качестве экспертов привлекались специалисты в области мониторинга цифровой экономики и статистики информационного общества, а также специалисты-социологи.

Перед специалистами в предметной области мониторинга ставилась задача оценить возможную избыточность или нерелевантность вопросов анкеты; указать на существенные аспекты предметной области, которые не нашли отражения в анкете; указать, какие формулировки вопросов следует изменить с точки зрения корректного использования принятой терминологии; дать экспертную оценку распространенности среди различных предприятий и организаций тех аспектов работы с большими данными, которые отражены в оцениваемой анкете.

Перед специалистами-социологами была поставлена задача оценить корректность и понятность для респондентов формулировок вопросов анкет, вариантов ответов и пояснений к ним; логику построения анкеты (последовательность вопросов, логичность и корректность переходов); полноту и корректность преамбулы к анкете; пригодность анкеты для различных методов опроса (личное интервью, телефонный опрос, онлайн-опрос).

4 Пилотное исследование и его результаты

Как и в других направлениях проведения мониторинга развития и использования технологий работы с большими данными, реализуемых в рамках проекта BD4DE (см. [2]), в настоящей работе для пилотирования (апробации разработанной методологии) были проведены кабинетные и экспертные исследования и оценки, причем основным использованным методом был метод экспертных оценок. Экспертная процедура, использованная для пилотажа, позволила протестировать некоторые аспекты методологии предлагаемого мониторинга и включала две стадии.

На первой стадии проводилась экспертная оценка концептуальной схемы и системы показателей для мониторинга. Для экспертной оценки привлекались специалисты в области мониторинга цифровой экономики и статистики информационного общества, имеющие опыт разработки систем показателей и инструментария для обследований, а также хорошо знакомые с имеющимися и потенциальными источниками данных. На этой стадии пилотажа оценивались полнота и адекватность разработанной концептуальной схемы, релевантность показателей целям исследования и, кроме того, источники и инструменты, выбранные для измерения показателей.

На второй стадии пилотажа доработанный инструментарий (анкеты для обследования организаций, экспертные анкеты или их разделы) был предложен для экспертной оценки двум группам экспертов:

- специалистам-социологам с опытом и квалификацией для разработки анкет для опроса организаций, а также экспертных анкет и экспертных процедур;
- специалистам в области финансирования развития и использования технологий работы с большими данными.

По результатам реализации первой стадии пилотной реализации был сделан вывод о необходимости уточнения состава и степени детализации показателей в части механизмов финансирования развития и использования технологий работы с большими данными без принципиальных изменений концептуальной схемы мониторинга. Неизменным остался набор показателей для оценки системы управления развитием и использованием технологий работы с большими данными. Одновременно набор показателей для мониторинга механизмов финансирования развития и использования технологий работы с большими данными, предложенный на начальной стадии концептуализации, было рекомендовано объединить в один комплексный показатель «Механизмы финансирования развития и использования технологий работы с большими данными». Он характеризует не только источники финансирования, но и обеспеченность финансированием как процессов разработки, так и процессов использования технологий работы с большими данными. Названный комплексный показатель был включен в концептуальную схему (см. рис. 1). Также по результатам первой стадии пилотной реализации были доработаны соответствующие анкеты для экспертных опросов по вопросам оценки системы управления и механизмов финансирования развития и использования технологий работы с большими данными.

На второй стадии пилотной реализации формируемой системы мониторинга экспертом-социологом были сделаны предложения по унификации всех возможных показателей разрезов, учитывающих обследуемые отрасли экономики, социальной сферы и системы государственного управления. В результате были доработаны соответствующие разделы анкеты для экспертного опроса, причем для нескольких показателей были введены дополнительные разрезы.

Таким образом, проведение пилотажа позволило не только протестировать и доработать основные составные части методологии мониторинга механизмов финансирования развития и использования технологий работы с большими данными (концептуальную схему, системы показателей, источники, инструментарий), но и получить обобщенную оценку ситуации, сложившейся в рассматриваемой предметной области, причем в различных ее аспектах.

Заключение

Сегодня в мире общепризнана необходимость концентрации ресурсов на наиболее важных направлениях развития и применении технологий, обладающих наибольшей эффективностью использования. Сказанное в полной мере относится к технологиям работы с большими данными. Для концентрации ресурсов на национальном уровне нужно построить такую систему управления,

которая сможет обеспечить целенаправленное воздействие на процессы развития и использования технологий. Эффективное управление позволяет организовать разработку и реализацию политики и стратегии научно-технологического развития, в том числе на основе технологий работы с большими данными. Задача формирования такой системы управления, а также связанных с ней механизмов финансирования развития и использования технологий работы с большими данными стоит сегодня перед Россией. Решение названной задачи без постоянного мониторинга соответствующей предметной области, очевидно, невозможно.

В настоящей работе представлены компоненты концептуальной схемы мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации, относящиеся к системе управления и механизмам финансирования работы с большими данными. Применимость предлагаемой концептуальной схемы и вошедших в нее показателей мониторинга доказана в рамках пилотной реализации с учетом рекомендованных доработок.

По результатам пилотной реализации системы мониторинга системы управления и механизмов финансирования развития и использования технологий работы с большими данными можно сделать несколько общих выводов:

- разработанная методология, скорректированная в ходе пилотажа, реализуема и готова для проведения полномасштабных исследований;
- масштабируемость разработанной методологии позволяет легко адаптировать ее к возможностям бюджета мониторинга и запросам потребителей – детализировать или укрупнять отраслевую структуру для представительного опроса предприятий, проводить мониторинг в разрезе субъектов Российской Федерации, страны в целом и т. д.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. World Bank. 2021. Unraveling Data's Gordian Knot: Enablers and Safeguards for Trusted Data Sharing in the New Economy. World Bank, Washington, DC. URL: <https://openknowledge.worldbank.org/handle/10986/35119> (дата обращения 01.06.2021)
2. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
3. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0> (дата обращения 01.10.2021)
4. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government, OECD Public Governance Policy Papers, No. 02, OECD Publishing, Paris. <https://doi.org/10.1787/f64fed2a-en> (дата обращения 01.06.2021)
5. Federal Data Strategy. 2020 Action Plan. URL: <https://strategy.data.gov/action-plan/> (дата обращения 01.06.2021)
6. Federal Data Strategy. URL: <https://strategy.data.gov/> (дата обращения 01.06.2021)
7. Постановление Правительства Российской Федерации от 02.03.2019 № 234 «О системе управления реализацией национальной программы «Цифровая экономика Российской Федерации». URL: <http://government.ru/docs/35964/> (дата обращения 01.06.2021)
8. Автономная некоммерческая организация «Цифровая экономика». URL: <https://www.data-economy.ru> (дата обращения 01.06.2021)

9. Ассоциация больших данных. Стратегия развития рынка больших данных до 2024 года. URL: <https://rubda.ru/wp-content/uploads/2020/03/strategiya-bolshih-dannyh-srednyaya.pdf> (дата обращения 01.06.2021)
10. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. 166 с. URL: <http://DECA.iis.ru> (дата обращения 01.06.2021)

BIG DATA GOVERNANCE AND FINANCIAL SUPPORT

Elizarov, Alexander Mikhailovich

*Doctor of physical and mathematical sciences, professor
Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University,
Department of software engineering, professor
Kazan, Russian Federation
amelizarov@gmail.com*

Padzhev, Valentin Valentinovich

*Institute for the Development of the Information Society (IIS), head of Directorate of legal programs
Moscow, Russian Federation
vpadzhev@iis.ru*

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Abstract

A framework and a set of indicators developed for monitoring the management system and financing mechanisms for working with big data are presented. The framework includes two components covering the very management system for the development and the use of big data technologies, as well as the mechanisms for financing these works. The availability and effectiveness of the management system for the development and use of big data technologies, the participation of stakeholders in these processes; the presence and condition of the monitoring system and its compliance with international systems; digital leadership and international cooperation in the field of big data ; sources, volumes and models of financing in this area; incentivizing and securing funding for big data activities are determined as main indicators. Methods for measuring monitoring indicators of the considered subject area are proposed.

Keywords

big data; big data technologies; big data governance; use of big data financing mechanisms; stakeholder participation; digital leadership; use of big data financial support; Big Data for Digital Economy; BD4DE

References

1. World Bank. 2021. Unraveling Data's Gordian Knot: Enablers and Safeguards for Trusted Data Sharing in the New Economy. World Bank, Washington, DC. URL: <https://openknowledge.worldbank.org/handle/10986/35119> (accessed 01.06.2021)
2. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologij raboty s bol'shimi dannymi // Informacionnoe obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
3. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0> (accessed 01.10.2021)
4. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government, OECD Public Governance Policy Papers, No. 02, OECD Publishing, Paris. <https://doi.org/10.1787/f64fed2a-en> (accessed 01.06.2021)
5. Federal Data Strategy. 2020 Action Plan. URL: <https://strategy.data.gov/action-plan/> (accessed 01.06.2021)
6. Federal Data Strategy. URL: <https://strategy.data.gov/> (accessed 01.06.2021)
7. Postanovlenie Pravitel'stva Rossijskoj Federacii ot 02.03.2019 № 234 «O sisteme upravleniya realizaciej nacional'noj programmy «Cifrovaya ekonomiki Rossijskoj Federacii». URL: <http://government.ru/docs/35964/> (accessed 01.06.2021)

8. Avtonomnaya nekommercheskaya organizatsiya «Cifrovaya ekonomika». URL: <https://www.data-economy.ru> (accessed 01.06.2021)
9. Associatsiya bol'shih dannyh. Strategiya razvitiya rynka bol'shih dannyh do 2024 goda. https://rubda.ru/wp-content/uploads/2020/03/strategiya-bolshih-dannyh_srednyaya.pdf (accessed 01.06.2021)
10. Analiz tekushchego sostoyaniya razvitiya cifrovoj ekonomiki v Rossii. M.: Institut razvitiya informacionnogo obshchestva, 2018. 166 s. URL: <http://DECA.iis.ru> (accessed 01.06.2021)

Информационное общество: политика и факторы развития

ЧЕЛОВЕЧЕСКИЙ КАПИТАЛ ДЛЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Янышен Анна Андреевна

*Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития,
младший научный сотрудник
Петрозаводск, Российская Федерация
annichekh@gmail.com*

Аннотация

Представлены концептуальная схема и набор показателей, разработанные для мониторинга человеческого капитала как фактора, влияющего на производство, использование и воздействие технологий работы с большими данными. В концептуальную схему включены показатели, характеризующие как наличный человеческий капитал для работы с большими данными, так и системы его производства. Предложены методы измерения показателей мониторинга человеческого капитала и приведены результаты измерения доступных показателей.

Ключевые слова

большие данные, технологии работы с большими данными; человеческий капитал; наличный человеческий капитал; система воспроизводства человеческого капитала; компетенции в области технологий работы с большими данными; Big Data for Digital Economy; BD4DE

Введение

Цифровая трансформация экономики, социальной сферы и государственного управления влечет за собой увеличение спроса на работников, обладающих компетенциями в области цифровых технологий – как на специалистов в области информационно-коммуникационных технологий (ИКТ), так и на других работников (включая руководителей), умеющих использовать цифровые технологии в своей профессиональной деятельности. Одним из ключевых драйверов цифровой трансформации являются технологии работы с большими данными. С увеличением объемов доступных данных и возможностей по извлечению пользы из их анализа растет значимость навыков использования различных инструментов для сбора, обработки, хранения и анализа больших данных практически для всех профессиональных групп и во всех сферах деятельности. В этой ситуации становится актуальной задача мониторинга и оценки человеческого капитала для работы с большими данными в стране.

В статье представлена методология мониторинга человеческого капитала как фактора, влияющего на производство, использование и воздействие технологий хранения и анализа больших данных, предлагается система показателей, позволяющая определить уровень наличного в стране человеческого капитала и оценить систему его воспроизводства, а также приведены результаты измерения доступных для расчета показателей.

© Шапошник С.Б., Янышен А.А., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_66

1 Определение предметной области

В руководстве по измерению человеческого капитала Европейской экономической комиссии ООН [1] прослеживаются истоки концепции человеческого капитала начиная с работ Адама Смита, который еще в 18 веке писал о важности «приобретенных и полезных способностей всех жителей или членов общества». Однажды приобретенные человеком, эти способности становятся «капиталом, закрепленным и как бы реализованным в его личности» [2].

Систематические исследования человеческого капитала стали проводиться с начала 1960-х, тогда же в научной оборот было введено и само понятие «человеческий капитал» («human capital»). Это понятие использовалось и разрабатывалось в рамках экономической науки; считалось, что концепция человеческого капитала позволяет объяснить большую разницу между увеличением объема производства в стране и увеличением объема вкладываемых в производство традиционных ресурсов (земля, рабочая сила и капитал) [3]. Основоположниками современной теории человеческого капитала считаются экономисты, нобелевские лауреаты Т. Шульц и Г. Беккер. Под человеческим капиталом в их трудах понимается совокупность приобретенных знаний, навыков, опыта и способностей, которые влияют на экономическую продуктивность человека и повышают его возможности на рынке труда [3-5].

За прошедшее время было предложено много определений человеческого капитала, но в основном концепция развивалась в русле экономического подхода, – изучалось воздействие на экономические показатели инвестиций в человеческий капитал. Вместе с тем, в последнее время в целом ряде работ предлагались определения человеческого капитала, расширенные по двум направлениям: включение в это понятие большего числа атрибутов человека и более широкого спектра эффектов от развития человеческого капитала. Так, если в отчете ОЭСР 1998 г. человеческий капитал определялся как «присущие людям знания, навыки, компетенции и другие атрибуты, которые имеют отношение к экономической деятельности» [6], то в более позднем отчете 2001 г. человеческий капитал определялся как «знания, навыки, компетенции и атрибуты, воплощенные в людях, которые способствуют созданию личного, социального и экономического благополучия» [7]. Близко по смыслу определение, используемое Всемирным банком в проекте измерения человеческого капитала: «Человеческий капитал включает в себя знания, навыки и здоровье, в которые люди инвестируют и которые накапливают на протяжении всей своей жизни, что позволяет им реализовать свой потенциал в качестве продуктивных членов общества» [8]. Более развернутое определение дает Британская энциклопедия: «Человеческий капитал, – это нематериальные коллективные ресурсы, которыми обладают отдельные лица и группы в составе данного населения. Эти ресурсы включают в себя все знания, таланты, навыки, способности, опыт, интеллект, образование, суждения и мудрость, которыми обладают индивидуально и коллективно, совокупная сумма которых представляет собой форму богатства, доступного странам и организациям для достижения их целей» [9].

В рамках данной предметной области мониторинга – человеческий капитал как фактор развития и использования технологий работы с большими данными – под человеческим капиталом мы будем понимать приобретенные в течение жизни знания, навыки, опыт и способности, которыми обладают отдельные лица и группы в населении страны, совокупная сумма которых представляет собой форму богатства, необходимого для достижения целей развития и использования технологий работы с большими данными в различных сферах деятельности.

Основной проблемой операционализации этого определения и измерения показателей человеческого капитала в сфере технологий работы с большими данными является то обстоятельство, что действующие классификаторы занятий, а также классификаторы направлений и специальностей подготовки, по которым собирается статистика занятых в экономике специалистов и их подготовки в учреждениях высшего и среднего профессионального образования, не обладают достаточной степенью детализации, чтобы выделить занятия и специальности, непосредственно связанные с работой с большими данными.

2 Обзор источников и литературы

2.1 Обзор подходов, концептуальных схем и показателей, разработанных международными организациями для мониторинга человеческого капитала как фактора развития цифровой экономики

Международные стандарты и системы мониторинга развития информационного общества и цифровой экономики, а также и действующие классификаторы пока не адаптированы для целей измерения такой достаточно специальной предметной области как человеческий капитал для развития и использования технологий работы с большими данными. Вместе с тем, выстраивая систему мониторинга этой предметной области, необходимо учитывать и использовать накопленный опыт и подходы, разработанные для измерения более широкой области – человеческого капитала для развития и использования цифровых технологий.

Важным инструментом измерения наличного человеческого капитала являются классификации занятий, на основе которых проводятся обследования рабочей силы с точки зрения профессиональных занятий и квалификации. Первая версия международной стандартной классификации занятий была принята Международной организацией труда в 1949 г. и несколько раз существенно обновлялась (примерно раз в 20 лет). С интересующей нас точки зрения – мониторинга наличия и квалификации ИКТ-специалистов – этапной является Международная стандартная классификация занятий 2008 г. (МСКЗ-08 [11] – она действующая). В МСКЗ-08 с учетом быстрого технического прогресса в этой сфере были существенно обновлены и расширены классификационные группы для этих занятий, что позволило, в частности, идентифицировать в рамках МСКЗ-08 профессионалов-специалистов и среднетехнический персонал в области ИКТ. Предложенная в МСКЗ-08 тематическая группа для занятий в области ИКТ включает две подгруппы 2-го уровня (25 и 35) и несколько начальных групп (4-го уровня). На основе МСКЗ-08 был разработан Общероссийский классификатор занятий, введенный в действие в 2014 г. [12]; данные по ИКТ-специалистам в экономике страны публикуется ежегодно, но номенклатура занятий в этой сфере недостаточно детализирована и не позволяет выделять специалистов в области работы с большими данными – есть только группа специалистов по базам данных.

Международные стандарты измерения подготовки ИКТ-специалистов в системе образования были представлены, в частности, в Руководстве по измерению использования ИКТ в образовании ЮНЕСКО 2008 г. [13], в котором были предложены 3 показателя, связанных с профессиональной подготовкой специалистов в сфере ИКТ (обучение, выпуск и гендерное соотношение выпуска ИКТ-специалистов). Для расчета этих показателей в руководстве предложена группировка связанных с ИКТ областей образования и профессиональной подготовки из классификатора Евростата 1999 г. В опубликованной Институтом статистики ЮНЕСКО в 2015 году Международной стандартной классификации областей образования и профессиональной подготовки 2013 (МСКО-О 2013) [14] выделен специальный раздел высокого уровня «Информационные и коммуникационные технологии» (06 Information and Communication Technologies), в соответствии с которым в настоящее время собирается статистика подготовки ИКТ-специалистов в странах ЕС и других странах, использующих эту классификацию. В России для действующих в стране перечня профессий и специальностей среднего профессионального образования и перечня специальностей и направлений подготовки высшего профессионального образования, по которым собирается статистика профессионального образования, группировки для ИКТ-специалистов официально не утверждены, но разрабатывались, в частности, для Индекса готовности регионов России к информационному обществу (для высшего образования), впервые опубликованного в 2005 г. [15], а также такие группировки (для среднего и высшего профессионального образования и для аспирантуры) разработаны и используются в ежегодном статистическом справочнике «Индикаторы цифровой экономики» [16]. Отметим, что и в международных, и в отечественных классификаторах для системы образования специалисты в области технологий работы с большими данными не выделены, поэтому данные по их подготовке недоступны. Вместе с тем, в рамках более широких специальностей и направлений подготовки российские вузы формируют программы подготовки (они указываются в выдаваемых дипломах и публикуются на сайтах вузов), которые более детализированы, и на этом уровне уже можно выделять программы, относящиеся к работе с большими данными. Названия этих программ пока не стандартизованы и статистика по ним не собирается, но, тем не менее, это открывает возможности построения показателей для мониторинга человеческого капитала в области технологий работы с большими данными.

Вопросы измерения человеческого капитала нашли отражение также в международных системах мониторинга и в стандартах измерения использования цифровых технологий в бизнесе и других сферах деятельности.

С исторически первых методологий мониторинга и оценки уровня использования информационно-коммуникационных технологий – оценок электронной готовности (e-readiness assessment) – человеческий капитал и ассоциированные с ним темы (навыки и человеческие ресурсы, программы обучения в области ИКТ и др.) входили в число предметных областей, относящихся к факторам развития и использования цифровых технологий [17].

В первом международном стандарте статистического наблюдения за использованием ИКТ бизнесом, разработанном ОЭСР, в модельную анкету для бизнеса 2001 г. [18] вошло несколько вопросов, связанных с человеческом капиталом:

- в составе барьеров, мешающих использованию ИКТ, были выделены три позиции, связанные с человеческим капиталом: (1) низкий уровень ИКТ-навыков персонала, (2) трудности с наймом ИКТ-специалистов, (3) нежелание работников использовать ИКТ;
- в качестве прокси-показателей ИКТ-навыков можно также рассматривать запрашиваемую долю персонала, использующего компьютер и компьютер, подключенный к интернету на регулярной основе.

В 2005 г. была опубликована новая версия модельного обследования использования ИКТ бизнесом ОЭСР [19], в которой сохранились вопросы об использовании персоналом на регулярной основе компьютера и интернета, а вопрос про барьеры стал относиться к использованию электронной коммерции, но с человеческим капиталом был связан только один обобщенный вариант барьеров: нехватка квалифицированных сотрудников для разработки, обслуживания или использования требуемой технологии.

В последнюю версию модельного обследования бизнеса по вопросам использования ИКТ ОЭСР, опубликованную в 2015 г. [20], был включен специальный модуль (ICT Skills), связанный с человеческим капиталом, в котором был представлен более широкий спектр предметных областей и показателей человеческого капитала:

- наличие (занятость) ИКТ-специалистов:
 - o доля предприятий, на которых работают ИКТ-специалисты;
 - o доля ИКТ-специалистов (% от числа занятых);
- обучение работников в сфере ИКТ:
 - o доля предприятий, обучающих ИКТ-специалистов;
 - o доля предприятий, обучающих персонал ИКТ-навыкам;
- найм ИКТ-специалистов:
 - o доля предприятий, предлагающих вакансии;
 - o доля предприятий, испытывающих трудности при найме ИКТ-специалистов;
 - o доля предприятий, испытывающих различные трудности при найме ИКТ-специалистов (отсутствие у кандидатов навыков в области ИКТ – технических, управленческих, связанных с интеграцией ИКТ в бизнес, а также слишком высокие требования к заработной плате);
- функции ИКТ, выполняемые собственными силами или внешними специалистами:
 - o доля предприятий, выполняющих различные функции, связанные с ИКТ, собственными силами;
 - o доля предприятий, выполняющих различные функции, связанные с ИКТ с привлечением внешних специалистов (отдельно – зарубежных).

В международных системах мониторинга развития цифровой экономики и общества, основанных на построении композитных индексов, человеческий капитал как фактор цифрового развития присутствовал с самого начала.

В индексе развития ИКТ (ICT Development Index, IDI), публикуемом Международным союзом электросвязи (МСЭ), традиционно одним из трех составляющих его подындеков является подындекс ИКТ-навыков, который характеризует способность использовать ИКТ эффективно [21]. Первоначально этот подындекс оценивался только на основе общих показателей человеческого капитала – грамотность взрослого населения (в более поздних выпусках – средняя продолжительность обучения), вовлеченность молодежи в среднее и в высшее образование. Это было связано с тем, что специфичные для ИКТ-навыков показатели были недоступны для

большинства стран и в качестве прокси использовались указанные показатели грамотности и образования. В 2017 г. с учетом разработанного МСЭ руководства по опросу населения и домохозяйств в подындекс навыков было предложено добавить композитный показатель доли населения, имеющего ИКТ-навыки, рассчитываемый на основе показателей совершения различных операций с использованием компьютера в течении 3-х месяцев перед опросом [22, 23]. Показатель был использован в отчете по измерению ИКТ 2018 г. [24, 25], но от его включения в индекс развития ИКТ впоследствии отказались из-за недоступности данных для слишком большого числа входящих в рейтинг стран [26].

В индексе развития электронного правительства ООН (E-Government Development Index, EGDl) также одним из трех компонентов является подындекс человеческого капитала как фактора использования электронных правительственных сервисов [27]. Для расчета этого подындекса используются аналогичные общие показатели: грамотность взрослого населения, средняя и ожидаемая продолжительность обучения, вовлеченность в образование.

В последней версии авторитетного Индекса готовности к сетевому миру (Network Readiness Index, NRI), который теперь готовится Институтом Портуланс [28], представлено несколько показателей человеческого капитала: грамотность взрослого населения, вовлеченность в высшее образование, ИКТ-навыки (показатель, основанный на результатах опроса менеджеров предприятий, ежегодно проводимого Всемирным экономическим форумом, – средняя оценка по 7-балльной шкале степени в какой активное население страны обладает достаточными цифровыми навыками), а также два показателя на основе МСКЗ-08: доли в рабочей силе специалистов высшего и среднего уровня квалификации.

Более специализированные показатели представлены в размерности «человеческий капитал» европейского индекса развития цифровых экономики и общества (Digital Economy and Society Index, DESI [29]). В ЕС проводится развернутое статистическое наблюдение за цифровым развитием по единым стандартам (разработанным ОЭСР и Евростатом), что позволяет использовать для мониторинга более «прицельные» показатели. В DESI 2020 г. подындекс человеческого капитала рассчитывается на основе 6 показателей:

- 1) доля населения, имеющего как минимум базовый уровень цифровых навыков;
- 2) доля населения, имеющего цифровые навыки выше базового уровня;
- 3) доля населения, имеющего как минимум базовые навыки работы с программным обеспечением;
- 4) доля ИКТ-специалистов в занятом населении;
- 5) доля женщин ИКТ-специалистов в занятом населении;
- 6) доля ИКТ-специалистов в общем выпуске специалистов.

Первые 3 показателя являются композитными и рассчитываются на основе обследования населения по вопросам использования ИКТ. Методика измерения основана на концептуальной эталонной модели цифровых компетенций для граждан, разработанной Еврокомиссией [30]: для каждого компонента модели предложены операции, которые респондент выполнял с использованием компьютера и интернета за последние 3 месяца, и на этой основе формируются критерии отнесения цифровых навыков респондента к четырем уровням: отсутствие навыков, низкий уровень, базовый, выше базового [31].

Наиболее представительный и структурированный набор показателей для измерения человеческого капитала как фактора цифрового развития предложен в Национальном индексе развития цифровой экономики, который опубликован в 2018 г. для 32 стран (страны ЕС, Норвегия, Швейцария, Россия и Турция [32]). Подындекс человеческого капитала рассчитан там на основе трех групп показателей, характеризующих: (1) наличный в стране человеческий капитал; (2) масштабы и качество системы воспроизводства человеческого капитала; (3) способность страны удерживать, привлекать и использовать талантливые и квалифицированные кадры. Для измерения первой предметной области использовались показатели наличия и дефицита квалифицированных специалистов и специалистов в сфере ИКТ (на основе МСКЗ-08), а также показатели цифровых и смежных навыков населения: доля населения, имеющего базовый или выше уровень цифровых навыков (по методике Евростата), способность рабочей силы креативно мыслить и решать задачи в технологически насыщенной среде (из обследования ОЭСР навыков взрослого населения в рамках программы PIACC). Масштабы и качество производства человеческого капитала для цифровой экономики (в рамках системы образования, обучения сотрудников в организациях и переподготовки кадров службами занятости) оценивались на основе статистики образования и

использования ИКТ-бизнесом (выпуск ИКТ-специалистов, обучение персонала в сфере ИКТ и др.), результатов обследования достижений учащихся (программа ОЭСР PISA), данных опроса менеджеров предприятий Всемирного экономического форума (WEF Executive Opinion Survey) и опроса предприятий Всемирного банка (WB Enterprise Surveys) [32].

В последние годы стали разрабатываться страновые индексы готовности к использованию и уровня использования «сквозных» технологий. Для смежных большим данным технологий искусственного интеллекта опубликовано уже несколько индексов [33-35]. К проблемам построения общих индексов цифрового развития (недоступность гармонизированной статистики использования цифровых технологий для широкого круга стран) здесь добавляются проблемы, связанные с тем, что действующие классификаторы, стандарты статистического наблюдения и программы мониторинга цифровых технологий недостаточно детализированы, чтобы измерять показатели, непосредственно связанные со «сквозными» технологиями. В этой ситуации для измерения человеческого капитала для развития и использования этих технологий (а этот важный компонент присутствует во всех методиках построения индексов) используются как более общие традиционные показатели международной статистики, которые упоминались выше, так и специализированные показатели, основанные на новых источниках. Например, для измерения наличного человеческого капитала и его производства для технологий искусственного интеллекта (ИИ) используются данные различных глобальных цифровых платформ: специалисты по ИИ в профессиональной социальной сети LinkedIn; данные по библиотекам в области ИИ на крупнейшей платформе разработки открытого программного обеспечения GitHub; скачивание пакетов, связанных с искусственным интеллектом, языков программирования R и Python (косвенный признак наличия специалистов в сфере ИИ); статистика курсов по ИИ на платформе Coursera и т.д. [34].

В условиях отсутствия официальной статистики источником информации о человеческом капитале для сквозных технологий (в том числе технологий работы с большими данными) являются специальные доклады и проекты, которые основаны на собственных эмпирических исследованиях и в ряде случаев носят мониторинговый характер. В контексте данной работы важно упомянуть опыт двух исследований международных организаций.

В отчете Объединенного исследовательского центра (JRC) Европейской комиссии «Академическое предложение продвинутых цифровых навыков в 2019–2020 гг. Международное сравнение» [36] анализируются предлагаемые образовательными учреждениями программы обучения в четырех областях: искусственного интеллекта, высокопроизводительных вычислений, кибербезопасности и науки о данных (Data Science). Источником данных является глобальная платформа Studyportals [37], на которой представлены более чем 207 000 программ из 3 700 университетов в более чем 120 странах. Для того чтобы выделить указанные программы, для каждой специализированной области был разработан набор ключевых слов, на основе которых проводился поиск программ. Полученные данные анализировались в разрезе стран, более широких областей образования (в соответствии с МСКЗ-08, например «ИКТ», «Бизнес, управление и право») и уровней программ (бакалавриат, магистратура, краткосрочные курсы дополнительного образования). Характерно, что только 37% программ ЕС по Data Science относится к области образования «Информационно-коммуникационные технологии» и 11% к области «Инженерия, производство и строительство», а на область «Бизнес, управление и право» приходится 27%. Это отличает тематику Data Science от остальных технологических направлений и свидетельствует, по мнению авторов доклада, о востребованности навыков в области работы с большими данными в различных профессиональных областях [36].

В докладе ОЭСР «Человеческий капитал, лежащий в основе искусственного интеллекта: спрос на профессии и навыки в размещенных онлайн вакансиях» приведен анализ взаимосвязи набора навыков в размещенных работодателями США и Великобритании в интернете вакансиях, связанных с искусственным интеллектом [38]. Анализ проводился на основе данных о вакансиях онлайн-ресурса Burning Glass Technologies (BGT [39]). Здесь также востребованные навыки в области ИИ выделялись с использованием ранее разработанного списка ключевых слов, связанных с ИИ.

2.2 Обзор научных публикаций, посвященных исследованию человеческого капитала в области технологий работы с большими данными

Для поиска последних научных работ, посвященных исследованию человеческого капитала для работы с большими данными, был проведен анализ публикаций, индексируемых на междисциплинарной поисковой платформе Web of Science (WoS).

Для отбора научных публикаций первоначально был сформирован поисковый образ, состоящий из ключевых слов, относящихся к человеческому капиталу (Human capital, Competence, Skills), объединенных логическим оператором AND с ключевыми словами, характеризующими технологии работы с большими данными [17]. За временной период с 2016 по 2021 г. по этому запросу в поисковом поле «Тема» (поиск по названиям, ключевым словам и аннотациям), было получено 2 302 публикации из WoS Core Collection, включающих исследования человеческого капитала для работы с большими данными, но в основном посвященным другим темам, – в частности, изучению различных аспектов человеческого капитала с использованием больших данных.

С учетом этого, были использованы два более «прицельных» запроса. По первому («HUMAN CAPITAL» AND «BIG DATA») было найдено 80 научных публикаций за последние 5 лет, а по второму ("BIG DATA COMPETENCE*" OR "DATA SCIENCE COMPETENCE*" OR "BIG DATA SKILL*" OR "DATA SCIENCE SKILL*") 60 публикаций, из которых были выбраны 15 релевантных [40-54].

Статья [40] описывает выгоды, которые может получить организация от инвестиций в аналитику больших данных (Big Data Analytics, BDA). В исследовании проверяется, в какой степени розничный бизнес может извлечь выгоду из инвестиций в BDA, имея доступ к большей доле выпускников с компетенциями в области больших данных на региональном уровне, в отличие от организаций, расположенных в районах, где количества подготовленных специалистов недостаточно. Эмпирический анализ показывает положительное влияние инвестиций бизнеса в аналитику больших данных, которые отражаются на технической эффективности как отдельных ритейлеров, так и отрасли в целом. При этом обнаружено, что если незначительная доля местной рабочей силы с набором профессиональных аналитических навыков может привести к негативному влиянию инвестиций в аналитику больших данных на техническую эффективность, то в тех случаях, когда доля такой рабочей силы достаточно велика, наблюдается положительная взаимосвязь между технической эффективностью бизнеса и инвестициями в BDA.

В статье [41] отмечается рост интереса организаций к большим данным и анализу за последние несколько лет, что влечет за собой интенсивный спрос на сертифицированных специалистов в области больших данных и науки о данных, а также отмечается, что улучшение аналитических навыков сотрудников позволит руководителям организаций принимать более эффективные решения, что повысит общую производительность компании. В исследовании отображается существование несоответствия потребностей отраслей в специалистах по большим данным и навыками, которые предлагают учебные заведения, и предлагается набор навыков, которыми должны обладать современные выпускники. В статье [42] поднимается вопрос поиска квалифицированного персонала с продвинутыми навыками анализа данных, а также анализируется спрос на компетенции в области больших данных, превышающий спрос на компетенции в бизнес-аналитике в Китае. В исследовании [43] отмечается важность аналитического мышления, необходимого при обработке данных, составлении прогнозов и принятии эффективных и действенных решений. Аналитическое мышление определяется как фундаментальный навык при обработке данных больших размерностей.

При изучении результатов запроса были также выделены публикации, содержащие исследования о необходимости навыков работы с большими данными у работников, не связанных с ИКТ отраслей: [44] – исследование возможностей использования больших данных в практике финансового менеджмента и выявление соответствующих навыков, необходимых в финансовом управлении, [45] – исследование необходимости аналитических навыков и навыков работы с большими данными у работников сферы гостеприимства, [46] – исследование влияния когнитивных информационных технологий на управленческий учет и разработка модели навыков работы с большими данными у бухгалтера, [47] – исследование влияния уровня аналитических навыков в качестве основы для успешной карьеры в области анализа больших данных в банковском секторе, [48] – статья о профессиональных навыках, которые могут повлиять на развитие

сотрудников и о том, как организации могут адаптировать навыки своих сотрудников к большим данным в сфере здравоохранения.

В публикациях отмечается, что нарастающие темпы использования аналитики больших данных влекут за собой изменения в поиске компаниями кандидатов на вакантные должности: исследование [49] посвящено выделению типов должностных обязанностей и необходимых для работы навыков; в исследовании [50] отмечается важность междисциплинарных навыков (вычислительной техники, математики, статистики, машинного обучения и знаний в области бизнеса) помимо ключевых навыков в области работы с большими данными, что способствует повышению эффективности бизнеса; в исследовании [51] сформулирован базовый набор навыков и компетенций, которыми должен обладать каждый специалист по большим данным.

Среди отобранных публикаций есть целый ряд посвященных исследованию подготовки выпускников с навыками работы в области больших данных образовательными учреждениями. В статье [52] анализируются представления о навыках в области работы с большими данными специалистов по работе с данными и преподавателей университета, отмечается, что достижение консенсуса поможет менеджерам по подбору персонала и в разработке рекомендаций для составления учебных программ. В статье [53] приведен количественный и качественный анализ текущих программ в области науки о данных. В исследовании [54] говорится о нехватке специалистов в области аналитики больших данных, неспособности образовательных учреждений подготовить достаточное количество специалистов для удовлетворения постоянно растущего спроса бизнеса на специалистов с навыками БД, анализируются возможности разработки междисциплинарной учебной программы, которая обеспечит навыками анализа данных студентов других академических специальностей (не ИКТ), что поможет выпускать более гибких специалистов по анализу данных и сократить разрыв между спросом и предложением на этих специалистов.

2.3 Выводы

Подводя итоги, можно сказать, что в международных стандартах, системах мониторинга и исследованиях человеческого капитала как фактора развития и использования цифровых технологий, разработаны подходы и системы показателей, характеризующие различные аспекты и составляющие человеческого капитала: наличный человеческий капитал (базовые цифровые навыки населения, наличие ИКТ-специалистов), системы его производства (в рамках системы образования или обучения персонала на предприятиях), а также спрос на цифровые компетенции в экономике. В случае «сквозных» технологий, к которым относятся технологии работы с большими данными и которые не представлены в действующей системе классификаторов и стандартах статистического наблюдения, для измерения человеческого капитала как фактора их эффективного использования, международными организациями, аналитическими компаниями и исследователями предложены новые методы и источники информации, предполагающие проведение собственных обследований и позволяющие решить проблему дефицита статистической информации в этой области.

3 Концептуальная схема и показатели мониторинга

3.1 Концептуальная схема предметной области

Как показал приведенный выше анализ международных стандартов измерения, систем мониторинга и исследований человеческого капитала как важного условия развития и использования цифровых технологий, эта предметная область имеет многомерную структуру. Прежде всего, важно рассматривать ее с точки зрения наличия цифровых компетенций и систем их производства (в рамках системы образования и подготовки кадров). При этом в фокусе внимания должны быть как специалисты в области цифровых технологий, так и цифровые навыки у выпускников и работников других профессиональных групп. Одновременно важно обращать внимание на то, насколько наличный человеческий капитал и системы его производства удовлетворяют растущий спрос на соответствующие компетенции в экономике.

С учетом сказанного, для мониторинга человеческого капитала как фактора развития и использования технологий работы с большими данными был разработан следующий подход. Комплексное измерение человеческого капитала страны для работы с большими данными должно включать две основные области мониторинга: (1) наличный в стране человеческий капитал;

(2) масштабы и качество системы производства человеческого капитала в области технологий работы с большими данными. При этом в той и другой предметных областях мониторинга должны быть представлены: (1) показатели наличия и подготовки специалистов в области технологий работы с большими данными, а также наличия и формирования знаний и навыков в этой сфере других профессиональных групп; (2) показатели наличия и удовлетворенности спроса на соответствующие компетенции и программы обучения (см. рисунок 1).

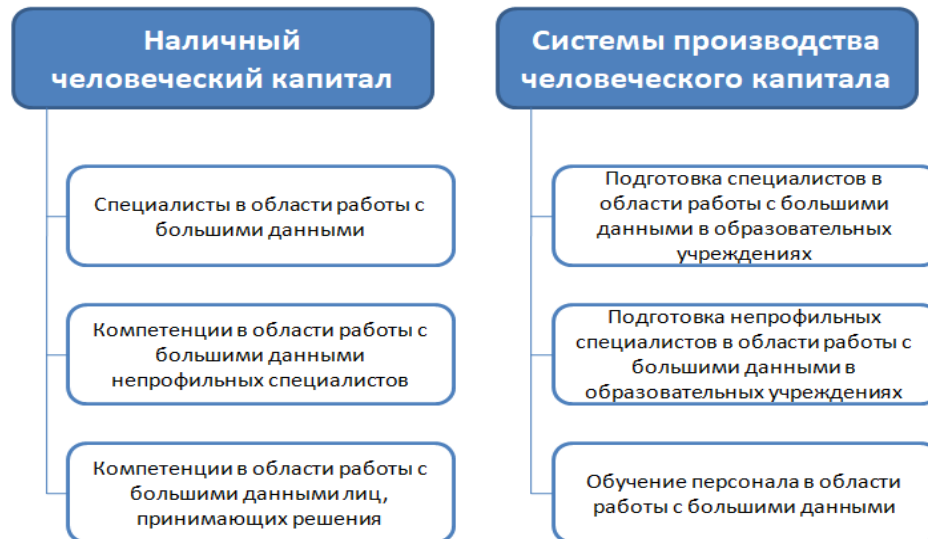


Рис. 1. Концептуальная схема мониторинга человеческого капитала как фактора развития и использования технологий хранения и анализа больших данных

Наличный человеческий капитал в области технологий работы с большими данными характеризуется показателями наличия специалистов, получивших профильное образование в данной сфере, а также показателями осведомленности и навыков работы с большими данными работников, не являющихся профильными специалистами, и цифровых лидеров – руководителей высшего звена организаций и предприятий.

Масштабы и качество системы воспроизводства человеческого капитала в области технологий работы с большими данными характеризуются показателями количества и доли образовательных учреждений, осуществляющих подготовку специалистов в области работы с большими данными и формированием необходимых навыков у выпускников непрофильных (не ИКТ) направлений подготовки, а также показателями масштабов обучения как специалистов в области работы с большими данными, так и остальных работников в организациях и предприятиях различных сфер деятельности.

Показатели удовлетворенности спроса организаций и предприятий в специалистах в области работы с большими данными является связующим элементом для двух областей мониторинга, отображая взаимосвязь спроса на кадры в области работы с большими данными с возможностями систем производства человеческого капитала этому спросу соответствовать, то есть подготавливать достаточное количество выпускников с необходимыми для рынка компетенциями.

3.2 Показатели наличного человеческого капитала

Движущей силой кадрового потенциала страны являются специалисты в области работы с большими данными, которые обладают соответствующим уровнем образования и занимаются разработкой и продвижением цифровых технологий. В то же время в современных реалиях для динамичного развития и использования цифровых технологий недостаточно одних только профессионалов в области ИКТ и больших данных; необходимым условием для развития цифровой экономики является формирование навыков работы с большими данными и у работников предприятий и организаций, не получивших специального образования.

Наряду с определением наличного уровня компетенций работников предприятий стоит отметить важность выявления уровня осведомленности о возможностях и навыков работы с большими данными у руководителей организаций и предприятий высокого уровня – цифровых лидеров. Развитие осведомленности и навыков в области работы с большими данными

руководителей должно послужить драйвером ускоренного внедрения технологий больших данных на предприятиях и формирования культуры их использования.

Измерение таких показателей позволит сформировать картину о совокупном человеческом капитале предприятий и организаций, включающем как лиц, принимающих решения, так и работников.

Помимо количества уже работающих специалистов и необходимых навыков у других работников предприятий и организаций, наличный в стране человеческий капитал в рассматриваемой сфере характеризуется также и динамикой спроса организаций и предприятий на кадровый потенциал с необходимыми для профессиональной деятельности компетенциями.

В соответствии со сказанным для мониторинга уровня наличного в стране человеческого капитала была разработана следующая система показателей.

- (НКБД-01-01) Число специалистов в области работы с большими данными на 10 000 занятых.

Данный показатель характеризует «вооруженность» экономики (и ее отраслей) специалистами в области работы с большими данными.

- (НКБД-01-02) Доля организаций, использующих специалистов в области работы с большими данными.

Показатель отображает масштабы привлечения специалистов в области работы с большими данными для цифровой поддержки и трансформации деловых процессов в организациях.

- (НКБД-01-03) Удельный вес заявленной потребности организаций в специалистах в области работы с большими данными (вакантных мест) в общем числе рабочих мест специалистов в области работы с большими данными.

Показатель используется для характеристики масштабов текущего спроса организаций на специалистов в области работы с большими данными.

- (НКБД-01-04) Доля организаций, испытывавших потребности в специалистах в области работы с большими данными.

Данный показатель отображает текущую потребность организаций в специалистах в области технологий работы с большими данными

- (НКБД-01-05) Доля организаций, испытывающих трудности при заполнении вакансий в области работы с большими данными за предыдущий календарный год.

Показатель характеризует организации, которые не смогли заполнить вакансии в области работы с большими данными. Среди причин, по которым вакансии не были заполнены, выделяются: неудовлетворенность соискателя заработной платой, недостаточный уровень компетенций соискателей; отсутствие соискателей на открытые должности.

- (НКБД-01-06) Доля организаций, в которых сотрудники осведомлены о возможностях технологии работы с большими данными.

Данный показатель отображает уровень осведомленности работников организаций, не имеющих профильного образования, о возможностях применения технологий работы с большими данными в своей профессиональной деятельности.

- (НКБД-01-07) Доля организаций, в которых сотрудники (неспециалисты) имеют необходимые компетенции в области работы с большими данными.

Показатель характеризует наличие необходимых навыков работы с большими данными среди работников организаций, не являющихся специалистами в данной сфере.

- (НКБД-01-08) Доля организаций, которые при заполнении вакансий (неспециалистов в области работы с большими данными) включают в квалификационные требования компетенции соискателей в области работы с большими данными.

Показатель используется для измерения характеристики спроса организаций на компетенции соискателей (неспециалистов) в области работы с большими данными.

- (НКБД-01-09) Доля организаций, в которых руководители высокого уровня осведомлены о возможностях работы с большими данными.

Фактором использования технологий работы с большими данными для совершенствования и трансформации бизнес-процессов организации является осведомленность о возможностях этих

технологий руководства организации. Для измерения этого параметра используется данный показатель.

- (НКБД-01-10) Доля организаций, в которых руководители высокого уровня используют аналитику больших данных в своей профессиональной деятельности (при принятии решений).

Показатель позволяет охарактеризовать наличие в организациях цифровых лидеров, владеющих соответствующими компетенциями для работы с большими данными.

Приведенный выше перечень показателей мониторинга и оценки уровня наличного в стране человеческого капитала рассматривается не только в масштабах всей страны, но и в разрезах по приоритетным сферам деятельности.

3.3 Показатели системы производства человеческого капитала

Основным элементом системы воспроизводства человеческого капитала является система образования, которую необходимо характеризовать как с точки зрения количества выпускаемых специалистов, так и с точки зрения качества получаемого ими образования; то есть необходима оценка того, насколько предоставляемые программы подготовки соответствуют спросу современных работодателей. Помимо подготовки специалистов в области технологий работы с большими данными необходимо оценивать формирование соответствующих компетенций у специалистов нетехнологической направленности, которые могут применять технологии работы с большими данными в своей профессиональной деятельности.

Системы воспроизводства человеческого капитала в стране представлены высшим и средним профессиональным образованием, а также дополнительным профессиональным образованием. Наряду с этим обучение и повышение квалификации в области работы с большими данными проводят также сами организации, использующие технологии, что также является важной частью системы производства человеческого капитала. Поэтому для оценки этой системы важным является мониторинг не только образовательных учреждений, но и оценка возможности обучения сотрудников в организациях (как специалистов в области технологий работы с большими данными, так и непрофильных специалистов).

С учетом сказанного для мониторинга системы производства человеческого капитала предлагается следующая система показателей.

- (ПКБД-02-01) Доля образовательных учреждений, выпускающих специалистов по программам подготовки в области работы с большими данными среди учреждений среднего профессионального образования, высшего образования, дополнительного профессионального образования.

Данный показатель позволит оценить масштаб подготовки специалистов в области работы с большими данными в образовательных учреждениях.

- (ПКБД-02-02) Доля образовательных учреждений, формирующих знания и навыки в области работы с большими данными выпускникам направлений подготовки, не связанным с ИКТ, среди среднего профессионального образования, высшего образования, дополнительного профессионального образования.

Показатель характеризует масштабы формирования навыков в области работы с большими данными выпускников не ИКТ-направлений подготовки.

- (ПКБД-02-03) Доля организаций, которые проводят обучение специалистов в области работы с большими данными.

Данный показатель отображает способность организаций проводить профильное обучение для сотрудников, являющихся специалистами в области работы с большими данными.

- (ПКБД-02-04) Доля организаций, которые проводят обучение персонала (неспециалистов) в области работы с большими данными.

Показатель позволяет измерить возможность организаций проводить обучение в области работы с большими данными сотрудников, не являющихся специалистами.

Аналогично показателям мониторинга и оценки уровня наличного в стране человеческого капитала показатели системы производства рассматриваются в разрезах по приоритетным сферам деятельности, а также по уровням образования.

4 Методология

Учитывая отмеченную выше проблему – действующие классификаторы занятий и формы статистического наблюдения не обладают достаточной степенью детализации, чтобы выделять специалистов и навыки в области работы с большими данными – оценка наличного в стране человеческого капитала возможна только путем проведения опроса организаций. На этапе пилотной реализации методологии мониторинга для получения информации, частично заменяющей данные статистики и опросов, было принято решение ввести дополнительные показатели, значения для которых можно получить путем проведения анализа открытой информации на платформе поиска работы hh.ru.

Для анализа вакансий в исследуемой предметной области мониторинга было выделено десять приоритетных сфер деятельности [55]:

- промышленность;
- сельское хозяйство;
- строительство;
- развитие городской среды;
- транспорт и логистика;
- энергетическая инфраструктура;
- финансовые услуги;
- здравоохранение;
- образование и наука;
- государственное управление.

В исследование в связи с востребованностью и массовым использованием работы с большими данными в данных сферах деятельности были включены две дополнительные отрасли:

- ИКТ-сектор;
- сектор контента и СМИ.

При анализе вакансий в перечисленных выше сферах деятельности было рассчитано два дополнительных показателя:

- (НКБД-01-10) Доля вакансий, требующих навыки работы с большими данными, в общем числе вакансий в сфере деятельности.

Расчет показателя проводился по формуле:

$$Дв = (Квнбд / Кв) * 100\%,$$

где

Дв – доля вакансий, требующих навыки работы с большими данными, в общем числе вакансий в сфере деятельности;

Квнбд – количество вакансий, требующих навыки работы с большими данными;

Кв – общее число вакансий в сфере деятельности.

- (НКБД-01-11) Доля вакансий, требующих навыки работы с большими данными, в числе всех вакансий в сфере деятельности, связанных с ИКТ.

Расчет показателя проводился по формуле:

$$Двикт = (Квнбд / Квикт) * 100\%,$$

где

Двикт – доля вакансий, требующих навыки работы с большими данными, в числе всех вакансий в сфере деятельности, связанных с ИКТ;

Квнбд – количество вакансий, требующих навыки работы с большими данными;

Кв – общее число вакансий в сфере деятельности.

Для характеристики системы производства кадрового потенциала на этапе пилотного запуска мониторинга для измерения был рассчитан показатель (ПКБД-02-01) в разрезе образовательных учреждений высшего образования. В соответствии с приказом Министерства труда и социальной защиты Российской Федерации от 06.07.2020 № 405н «Об утверждении профессионального стандарта «Специалист по большим данным»» в требования к специалисту по большим данным

входит получение высшего образования – бакалавриата по следующим специальностям из общероссийского классификатора специальностей по образованию (ОКСО):

- 01.03.02 Прикладная математика и информатика
- 02.03.01 Математика и компьютерные науки
- 02.03.02 Фундаментальная информатика и информационные технологии
- 09.03.01 Информатика и вычислительная техника
- 09.03.02 Информационные системы и технологии
- 09.03.03 Прикладная информатика
- 38.03.05 Бизнес-информатика

Для расчета показателя ПКБД-02-01 был проведен анализ учебных планов 125 вузов СЗФО (<https://monitoring.miccedu.ru/?m=vpo>). Расчет показателя проводился по формуле:

$$\text{Доусз} = (\text{Коубд} / \text{Коусз}) * 100\%,$$

где

Доусз – доля образовательных учреждений высшего образования СЗФО, выпускающих специалистов по программам подготовки в области работы с большими данными;

Коубд – количество образовательных учреждений высшего образования СЗФО, выпускающих специалистов по программам подготовки в области работы с большими данными;

Коусз – общее количество образовательных учреждений высшего образования СЗФО.

5 Результаты

5.1 Наличный человеческий капитал

Для характеристики наличного человеческого капитала для работы с большими данными был произведен расчет двух дополнительных показателей (НКБД-01-10) и (НКБД-01-11), основанных на анализе платформы hh.ru (см. рисунок 2).

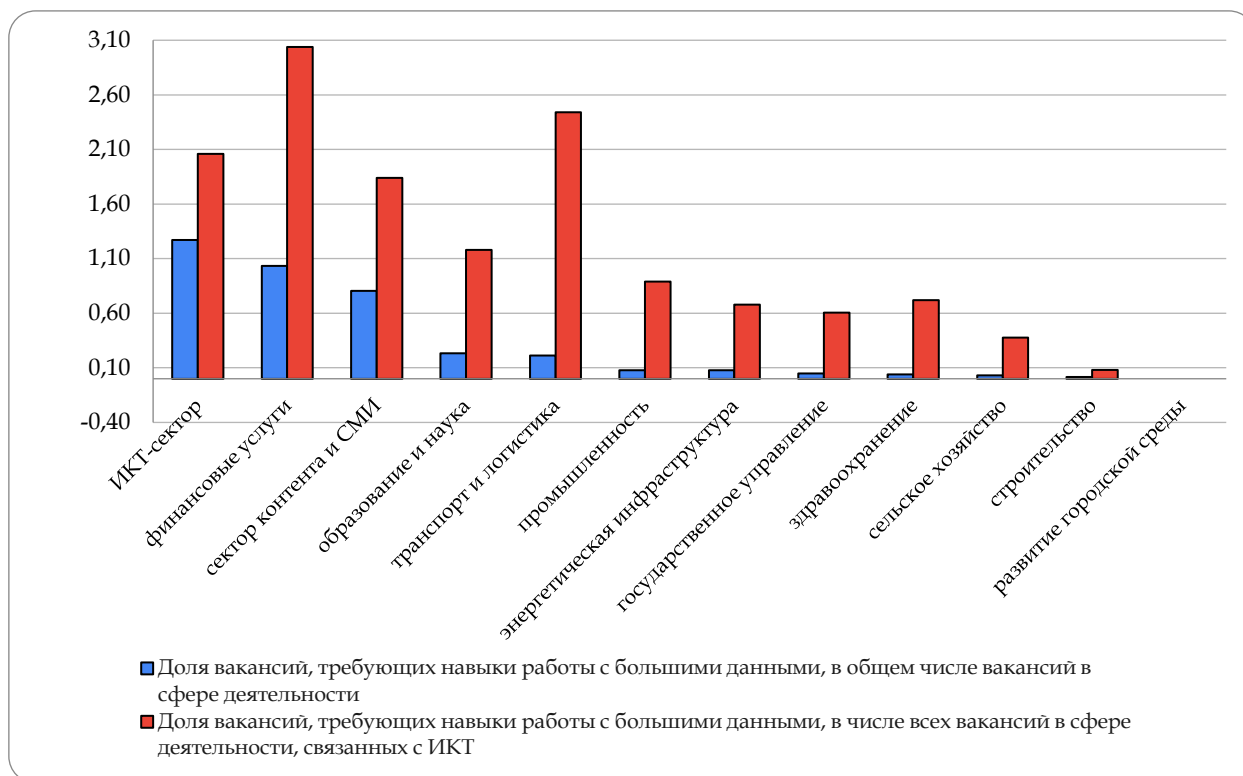


Рис.2. Доля вакансий, требующих навыки работы с большими данными, в общем числе вакансий и в числе всех вакансий в рассматриваемой сфере деятельности, связанных с ИКТ, май 2021 г.

Источник: данные платформы hh.ru, расчеты авторов

Полученные результаты демонстрируют, что наибольшую долю вакансии с навыками работы с большими данными от всех вакансий составляют в таких сферах деятельности, как ИКТ-сектор, финансовые услуги, сектор контента и СМИ. Что касается требований навыков в области работы с большими данными в вакансиях, связанных с профессиональными занятиями в области ИКТ (например, для ИКТ-специалистов), то наибольшую долю такие вакансии составляют в финансовом секторе, транспорте и логистике, ИКТ-секторе, секторе контента и СМИ и в образовании и науке, см. рисунок 2). Это является косвенным признаком нарастающей интенсивности использования технологий работы с большими данными в этих областях.

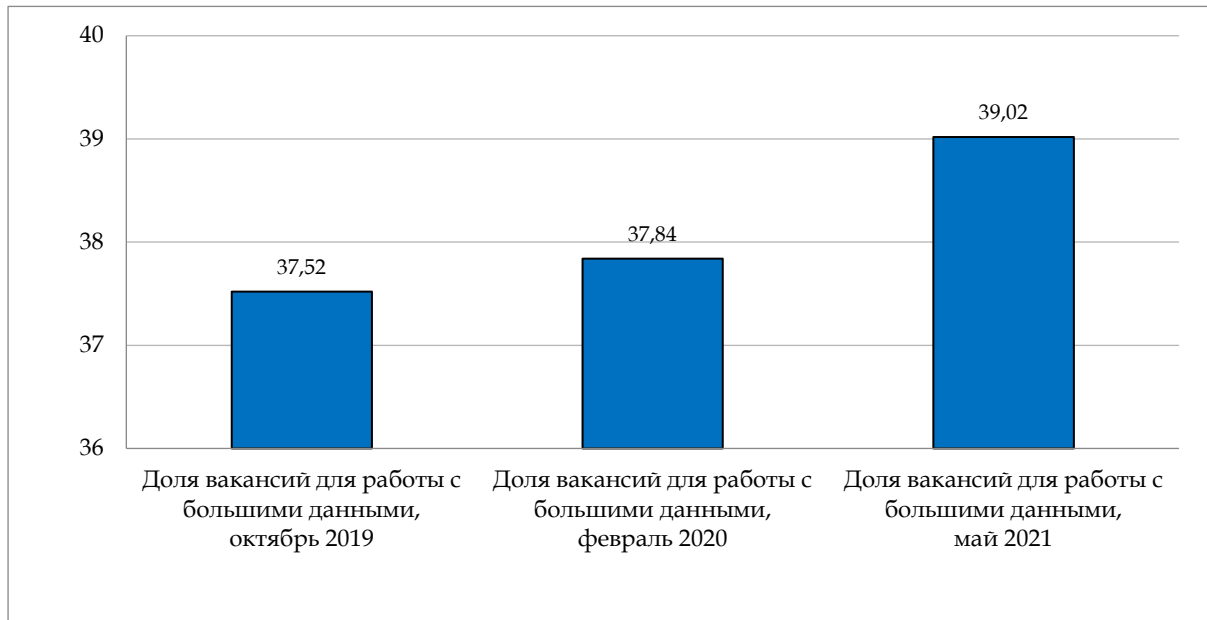


Рис.3. Доля вакансий для профессиональных занятий, не связанных с ИКТ, в общем числе вакансий, требующих компетенции в области работы с большими данными, 2019–2021 гг.

Источник: данные платформы hh.ru, расчеты авторов

Помимо данных по вакансиям в приоритетных сферах деятельности использовались данные платформы hh.ru о количестве вакансий, требующих навыков работы с большими данными, для различных видов профессиональных занятий (профессий) в разные периоды времени: за октябрь 2019 г., февраль 2020 г. и май 2021 г.

На рисунке 3 видна динамика положительного роста доли вакансий для профессиональных занятий, не связанных с ИКТ, в общем числе вакансий, связанных с большими данными. Полученные результаты позволяют сделать вывод о нарастающем спросе (с 37,5% до 39%) на навыки работы с большими данными у работников, не являющихся специалистами в области ИКТ. Иными словами, опережающими темпами идет рост спроса на компетенции в области работы с большими данными у различных профессиональных групп, не связанных напрямую с ИКТ, что свидетельствует о все более широком использовании технологий работы с большими данными профессионалами в различных областях.

5.2 Системы производства человеческого капитала

В данном разделе представлены результаты анализа систем производства человеческого капитала для работы с большими данными на основе измерения показателей мониторинга в ходе пилотного исследования в регионах Северо-Западного федерального округа (СЗФО).

В рамках пилотной реализации проекта для мониторинга систем производства человеческого капитала в области работы с большими данными был рассчитан показатель (ПКБД-02-01) в разрезе по образовательным учреждениям высшего образования СЗФО.

Для расчета показателя были выбраны высшие учебные заведения Северо-Западного федерального округа (<https://monitoring.miccedu.ru/?m=vpo>) и проведен анализ 125 образовательных организаций высшего образования, включая филиалы и вузы Министерства обороны (6 учебных заведений). Среди них было выделено 52 учебных заведения, в которых происходит подготовка специалистов по описанным выше специальностям, что составило 41,6%

среди всех заведений высшего образования в СЗФО. Далее был проведен анализ учебных планов, находящихся в открытом доступе на официальных веб-сайтах выбранных вузов, на предмет наличия дисциплин, характеризующих подготовку специалистов в области больших данных. Среди 52 учебных заведений, проводящих подготовку ИКТ-специалистов, было выявлено 40 учебных заведений, чьи учебные планы включали в себя такие дисциплины как «Введение в анализ больших данных», «Машинное обучение», «Анализ пространственных данных», «Анализ Big Data», «Технологии Big Data», «Хранение и обработка больших объемов данных», «Интеллектуальный анализ данных», «Технология анализа больших данных», «Решения на основе технологии Big Data» и пр.

Явным лидером по всем показателям (как по общему числу вузов, так и по числу вузов с направлениями подготовки в сфере ИКТ и подготовки специалистов в области работы с большими данными) ожидаемо выступает центр округа – город федерального значения Санкт-Петербург.

В качестве характеристики ситуации с системой производства человеческого капитала в области работы с большими данными в СЗФО был рассчитаны значения показателя (ПКБД-02-01) «Доля образовательных учреждений, выпускающих специалистов по программам подготовки в области работы с большими данными» в разрезе по всем учебным заведениям высшего образования и по учебным заведениям, осуществляющим подготовку специалистов в области ИКТ в соответствии со списком специальностей по ОКСО, перечисленным в разделе 4 (рисунки 4).

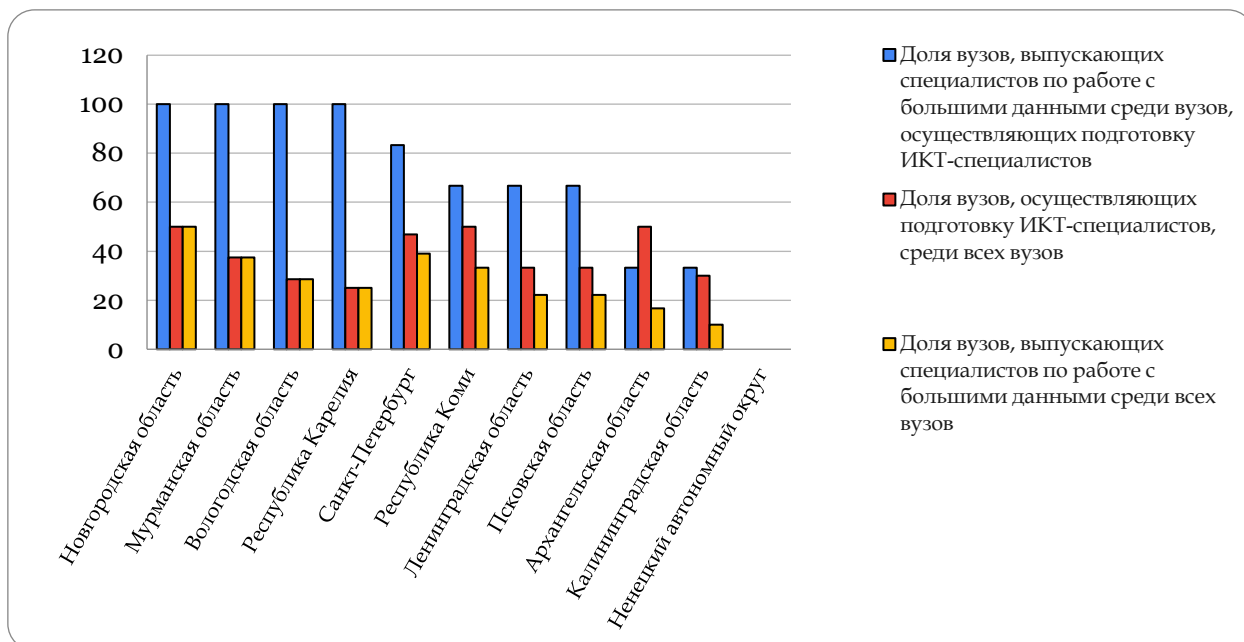


Рис. 4. Доля образовательных учреждений, выпускающих специалистов по программам подготовки в области работы с большими данными среди всех учебных заведений высшего образования и среди учебных заведений, осуществляющих подготовку специалистов в области ИКТ

В Новгородской, Мурманской, Вологодской областях и Республике Карелия подготовку специалистов в области работы с большими данными осуществляют все учебные заведения высшего образования, которые проводят подготовку ИКТ-специалистов. В Санкт-Петербурге соотношение таких вузов 83,3%, в Республике Коми, Ленинградской и Псковской областях – 66,7%, наименьшие показатели у Архангельской и Калининградской областей – 33,3%. В Ненецком АО на 2021 г. отсутствуют учреждения высшего образования.

При расчете доли учебных заведений высшего образования с подготовкой специалистов в области работы с большими данными среди всех вузов наилучшие результаты у Новгородской области – 50%, а также в Санкт-Петербурге и Мурманской области – 39% и 37,5% соответственно.

Абсолютным лидером Северо-Западного федерального округа является Национальный исследовательский университет ИТМО, который насчитывает 34 направления, связанных с подготовкой специалистов в области работы с большими данными.

Вузы-лидеры остальных субъектов СЗФО по количеству направлений подготовки специалистов в области работы с большими данными представлены в таблице 1.

Таблица 1. Распределение вузов субъектов СЗФО по числу направлений подготовки специалистов в области работы с большими данными

Субъект СЗФО	Наименование ВУЗа	Направления подготовки в области больших данных
Калининградская область	Балтийский федеральный университет имени Иммануила Канта	9
Псковская область	Псковский государственный университет	8
Республика Карелия	Петрозаводский государственный университет	7
Мурманская область	Мурманский государственный технический университет	6
Республика Коми	Сыктывкарский государственный университет им. Питирима Сорокина	6
Вологодская область	Вологодский государственный университет	4
Архангельская область	Северный (Арктический) Федеральный университет им. М.В.Ломоносова	3
Новгородская область	Новгородский государственный университет имени Ярослава Мудрого	3
Ленинградская область	Государственный институт экономики, финансов, права и технологий	1
Ленинградская область	Ивангородский гуманитарно-технический институт (филиал) Санкт-Петербургского государственного университета аэрокосмического приборостроения	1

Заключение

Международные стандарты статистического наблюдения и системы мониторинга развития и использования цифровых технологий, а также действующие классификаторы пока не адаптированы для целей измерения таких достаточно специальных предметных областей, как человеческий капитал для развития и использования «сквозных» технологий, к которым относятся технологии работы с большими данными. Вместе с тем спрос на мониторинговую информацию в этих областях весьма велик, учитывая значимость этих технологий для современных процессов цифрового развития различных сфер деятельности.

В этих условиях предложенные в данной работе концептуальная схема, показатели и методы измерения могут использоваться не только для задач мониторинга человеческого капитала в области работы с большими данными, но и являться образцом для решения аналогичных задач для других технологических направлений (использование искусственного интеллекта, аддитивных технологий, робототехники, квантовых технологий и др.).

Для целей формирования государственной системы мониторинга человеческого капитала для развития и использования технологий работы с большими данными (аналогично и для других «сквозных» технологий) актуальными являются следующие задачи:

- разработка моделей компетенций, профессиональных и образовательных стандартов в области технологий работы с большими данными;
- стандартизация названий программ подготовки в области технологий работы с большими данными и включение данных о подготовке специалистов по этим программам в собираемую статистику профессионально образования;
- детализация форм статистического наблюдения в сфере использования цифровых технологий для получения сведений о наличии и дефиците специалистов в области технологий работы с большими данными.

При отсутствии такой государственной системы статистического наблюдения основными источниками мониторинговой информации могут являться опросы организаций различных сфер деятельности, а также специальные исследования с использованием альтернативных источников данных (платформ для поиска работы, сайтов образовательных учреждений и др.) и различных методов их анализа. Дальнейшее развитие этих методов и расширение используемых источников данных остается актуальной задачей на ближайшие годы.

Проведенные в рамках пилотной реализации методологии мониторинга исследования демонстрируют расширяющийся спрос на компетенции в области технологий работы с большими данными не только для специалистов в сфере ИКТ, но и для широкого круга других профессиональных занятий. Учреждения профессионального образования стараются удовлетворить этот спрос и расширяют предложения программ подготовки и курсов в этой сфере.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. Guide on Measuring Human Capital. United Nations. 2016. URL: <https://unstats.un.org/unsd/nationalaccount/consultationDocs/HumanCapitalGuide.web.pdf> (дата обращения 01.06.2021)
2. Adam Smith An Inquiry into the Nature and Causes of the Wealth of Nations. 1776. (русский перевод: Адам Смит Исследование о природе и причинах богатства народов. – Москва, 2016. – 1056 с)
3. Schultz T. W. Capital Formation by Education. Journal of Political Economy, 1960, vol. 68, No. 6, pp. 571–583.
4. Schultz T. W. Investment in Human Capital. American Economic Review, 1961, vol. 51, No. 1, pp. 1–17.
5. Becker, Gary S. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. 1964. New York: Columbia University Press.
6. Human Capital Investment. An International Comparison. OECD, 1998. URL: <https://www.oecd-ilibrary.org/docserver/9789264162891-en.pdf?expires=1634023193&id=id&accname=guest&checksum=129BA138898AA8729161CE354777C4DB> (дата обращения 01.06.2021)
7. The Well-being of Nations. The Role of Human and Social Capital. OECD, 2001. URL: <https://www.oecd-ilibrary.org/docserver/9789264189515-en.pdf?expires=1634023630&id=id&accname=guest&checksum=422F778380B5AB356A774DC93897A759> (дата обращения 01.06.2021)
8. World Bank. The Human Capital Project: Frequently Asked Questions. URL: https://www.worldbank.org/en/publication/human-capital/brief/the-human-capital-project-frequently-asked-questions?cid=GGH_e_hcpexternal_en_ext#HCP2 (дата обращения 15.06.2021)
9. Encyclopedia Britannica. Human Capital. URL: <https://www.britannica.com/topic/human-capital>. (дата обращения 15.06.2021)
10. ISCO-88: International Standard Classification of Occupations. – Geneva, 1990.
11. International Standard Classification of Occupations: ISCO 08 / International Labour Office. -. Geneva: ILO, 2012. URL: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf (дата обращения 15.06.2021)

12. ОК 010–2014 (МСКЗ-08). Общероссийский классификатор занятий (принят и введен в действие Приказом Росстандарта от 12.12.2014 N 2020-ст). URL: <https://docs.cntd.ru/document/1200121893> (дата обращения 01.06.2021)
13. Guide to Measuring Information and Communication Technologies (ICT) in Education. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (дата обращения 01.06.2021)
14. International Standard Classification of Education: Fields of Education and Training 2013. Detailed field descriptions. UNESCO Institute for Statistics, Montreal, Quebec, 2015. URL: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-fields-of-education-and-training-2013-detailed-field-descriptions-2015-en.pdf> (дата обращения 15.06.2021)
15. Индекс готовности регионов России к информационному обществу. 2004–2005 / Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. – М.: Институт развития информационного общества, 2005. – 224 с. URL: <http://eRegion.ru> (дата обращения 01.06.2021)
16. Индикаторы цифровой экономики: 2021: статистический сборник / Г. И. Абдрахманова, К. О. Вишневыский, Л. М. Гохберг и др.; Нац. исслед. ун-т «Высшая школа экономики». – М.: НИУ ВШЭ, 2021. – 380 с.
17. Ершова Т.В., Хохлов Ю.Е., Шапошник С.Б. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
18. Measuring ICT Usage and Electronic Commerce in Enterprises: Proposal for a Model Questionnaire. OECD, Paris, 11-12 October 2001. URL: <https://www.oecd.org/digital/ieconomy/20628443.pdf> (дата обращения 01.06.2021)
19. ICT Use by Businesses: Revised OECD Model Survey. OECD, Paris, 2005. URL: <https://www.oecd.org/digital/ieconomy/35867672.pdf> (дата обращения 01.06.2021)
20. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (дата обращения 01.06.2021)
21. Measuring the Information Society Report 2017. Volume 1 – Geneva: International Telecommunication Union, 2017. – xi + 155 p.
22. Extraordinary Meeting of the Expert Group on ICT Household Indicators (EGH) and Expert Group on Telecommunication/ICT Indicators (EGTI). URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/events/eghegti2017/default.aspx> (дата обращения 01.06.2021)
23. The ICT Development Index (IDI): Methodology, indicators and definitions (as of February 2019). International Telecommunication Union, 2019. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/ITU_ICT%20Development%20Index.pdf (дата обращения 01.06.2021)
24. Measuring the Information Society Report 2018. Volume 1 – Geneva: International Telecommunication Union, 2018. – xi + 204 p.
25. Measuring the Information Society Report 2018. Volume 2 – Geneva: International Telecommunication Union, 2018. – viii + 244 p.
26. ICT Development Index 2020: A proposal. International Telecommunication Union, 2020. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/events/egti2020/IDI2020_BackgroundDocument_20200903.pdf (дата обращения 15.06.2021)
27. United Nations E-Government Survey 2020. - UN, New York, 2020. - 323 p. URL: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020> (дата обращения 15.06.2021)
28. Network Readiness Index 2020: Benchmarking the Future of the Network Economy. Portulance Institute, October 2020. – 317 p. URL: <https://networkreadinessindex.org/> (дата обращения 15.06.2021)
29. The Digital Economy and Society Index (DESI). URL: <https://ec.europa.eu/digital-single-market/en/desi> (дата обращения 15.06.2021)
30. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model. European Commission, 2016. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101254/jrc101254_digcomp%20

- [2.0%20the%20digital%20competence%20framework%20for%20citizens.%20update%20phase%201.pdf](#) (дата обращения 01.06.2021)
31. Individuals who have basic or above basic overall digital skills by sex. Eurostat Metadata. URL: https://ec.europa.eu/eurostat/cache/metadata/en/tepsr_sp410_esmsip2.htm. (дата обращения 01.06.2021)
 32. Национальный индекс развития цифровой экономики: Пилотная реализация. Госкорпорация «Росатом». Под ред. Т.В. Ершовой, Ю.Е. Хохлова, С.Б. Шапошника. 2018. – 92 с.
 33. Artificial Intelligence Index Report 2021. Stanford University. Human-Centered Artificial Intelligence Institute (HAI). 2021. – 222 p.
 34. Global AI Index. Tortoise Media. URL: <https://www.tortoisemedia.com/intelligence/global-ai/> (дата обращения 01.06.2021)
 35. Government AI Readiness Index 2020. Oxford Insights. 2020. – 144 p.
 36. Righi, R., López-Cobo, M., Alaveras, G., Samoili, S., Cardona, M., Vázquez-Prada Baillet, M., Ziemba, L.W., and De Prato, G., Academic offer of advanced digital skills in 2019-20. International comparison. Focus on Artificial Intelligence, High Performance Computing, Cybersecurity and Data Science, EUR 30351 EN, Publications Office of the European Union, Luxembourg, 2020. URL: <https://op.europa.eu/en/publication-detail/-/publication/fad44166-f249-11ea-991b-01aa75ed71a1> (дата обращения 01.06.2021)
 37. Study portal. URL: <https://studyportals.com/>. (дата обращения 15.06.2021)
 38. Lea Samek, Mariagrazia Squicciarini and Emile Sammeraat. The human capital behind AI: jobs and skills demand from online job postings. OECD, 2021. – 99 p. URL: https://www.oecd-ilibrary.org/science-and-technology/the-human-capital-behind-ai_2e278150-en (дата обращения 15.06.2021)
 39. *Burning Glass Technologies*. URL: <https://www.burning-glass.com/>. (дата обращения 15.06.2021)
 40. Sena, Vania & Ozdemir, Sena. (2020). Spillover effects of investment in big data analytics in B2B relationships: What is the role of human capital? *Industrial Marketing Management*. 86. 77-89. <https://doi.org/10.1016/j.indmarman.2019.05.016>.
 41. K. Rasheva-Yordanova, E. Iliev, V. Chantov, Analysis Of Missing Data Science Competence In It Sector”. Proceedings of EDULEARN18 Conference 2nd-4th July 2018, Palma, Mallorca, Spain IATED, ISBN: 978-84-09-02709-5, pp. 7399 – 7403, 201. <http://doi.org/10.21125/edulearn.2018.1731>
 42. Jun Wu, Honglei Shi, and Jiaping Yang, “Are big data talents different from business intelligence expertise? Evidence from text mining using job recruitment advertisements,” in 2017 International Conference on Service Systems and Service Management, Jun. 2017, pp. 1–6, ISSN:2161-1904. <https://doi.org/10.1109/ICSSSM.2017.7996289>
 43. K. Rasheva-Yordanova, E. Iliev, B. Nikolova, “Analytical Thinking As A Key Competence For Overcoming The Data Science Divide,” in Proceedings of EDULEARN18 Conference 2-4 July 2018, Palma, Mallorca, Spain, pp 7892-7898, 2018. <http://doi.org/10.21125/edulearn.2018.1833>
 44. Pilipczuk, O.; Cosenco, N.; Kosenko, O. Big Data: Challenges and opportunities in financial management. *Manag. Issues* 2019, 5, 9–23. <https://doi.org/10.7172/1644-9584.85.1>
 45. Mottier, E., Marshall, T. An Investigation into the Importance of Analytical Competencies: a Hospitality Industry Perspective. In: Proceedings of The 11th EuroMed Conference, 12-14 September, 2018. ISBN: 978-9963-711-67-3.
 46. Pilipczuk O. Toward Cognitive Management Accounting. *Sustainability* 2020, 12(12), 5108; <https://doi.org/10.3390/su12125108>
 47. Hakimi, A.A.A., Ali, M.S.M., Wahid, H. Big Data Skills Required for Successful Application Implementation in the Banking Sector, in Modeling, Design and Simulation of Systems Communications in Computer and Information Science, Vol. 752, 381-392 – August, 2017. https://doi.org/10.1007/978-981-10-6502-6_34
 48. Pesqueira, A., Sousa, M.J. & Rocha, Á. Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals. *J Med Syst* 44, 197 (2020). <https://doi.org/10.1007/s10916-020-01665-9>
 49. A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, “Human resources for Big Data professions: A systematic classification of job roles and required skill sets,” *Information Processing & Management*, vol. 54, no. 5, pp. 807–817, Sep. 2018. <https://doi.org/10.1016/j.ipm.2017.05.004>

50. Akhtar, P., Frynas, J.G., Mellahi, K. and Ullah, S. (2019), Big Data-Savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *Brit J Manage*, 30: 252-271.
<https://doi.org/10.1111/1467-8551.12333>
51. K. Rasheva-Yordanova, "Forming of Data Science Competence for Bridging the Digital Divide," in *The Future Education*, 2018
52. Catherine Chen & Haoqiang Jiang (2020) Important Skills for Data Scientists in China: Two Delphi Studies, *Journal of Computer Information Systems*, 60:3, 287-296,
<https://doi.org/10.1080/08874417.2018.1472047>
53. T. Wiktorski, Y. Demchenko, A. Belloum and A. Shirazi, "Quantitative and Qualitative Analysis of Current Data Science Programs from Perspective of Data Science Competence Groups and Framework," 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2016, pp. 633-638, <https://doi.org/10.1109/CloudCom.2016.0109>
54. J. C. Nwokeji, R. Stachel, T. Holmes, F. Aqlan, E. C. Udenze and R. Orji, "Panel: Addressing the Shortage of Big Data Skills with Inter-Disciplinary Big Data Curriculum," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-4, <https://doi.org/10.1109/FIE43999.2019.9028350>
55. Указ Президента Российской Федерации от 21 июля 2020 г. N 474 «О национальных целях развития Российской Федерации на период до 2030 года». URL:
<http://publication.pravo.gov.ru/Document/View/0001202007210012> (дата обращения 01.06.2021)

HUMAN CAPITAL FOR USE OF BIG DATA IN THE RUSSIAN FEDERATION

Shaposhnik, Sergei Borisovich

*Karelian Research Centre of the Russian Academy of Sciences, Department of multidisciplinary scientific research, Laboratory of digital technologies for regional development, senior researcher
Petrozavodsk, Russian Federation
sergei.shaposhnik@gmail.com*

Yanyshen, Anna Andreevna

*Karelian Research Centre of the Russian Academy of Sciences, Department of multidisciplinary scientific research, Laboratory of digital technologies for regional development, junior researcher
Petrozavodsk, Russian Federation
annichekh@gmail.com*

Abstract

The article presents a framework and a set of indicators developed to monitor human capital as a factor affecting the production, use and impact of big data storage and analysis technologies. The framework includes indicators characterizing both the available human capital for working with big data and its production systems. Methods of measuring human capital monitoring indicators are proposed and the results of measuring available indicators are presented.

Keywords

big data technologies; human capital; available human capital; production system of human capital; big data competencies; big data skills; Big Data for Digital Economy; BD4DE

References

1. Guide on Measuring Human Capital. United Nations. 2016. URL: <https://unstats.un.org/unsd/nationalaccount/consultationDocs/HumanCapitalGuide.web.pdf> (accessed on 01.06.2021)
2. Adam Smith An Inquiry into the Nature and Causes of the Wealth of Nations. 1776.
3. Schultz T. W. Capital Formation by Education. Journal of Political Economy, 1960, vol. 68, No. 6, pp. 571–583.
4. Schultz T. W. Investment in Human Capital. American Economic Review, 1961, vol. 51, No. 1, pp. 1-17.
5. Becker, Gary S. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. 1964. New York: Columbia University Press.
6. Human Capital Investment. An International Comparison. OECD, 1998. URL: <https://www.oecd-ilibrary.org/docserver/9789264162891-en.pdf?expires=1634023193&id=id&accname=guest&checksum=129BA138898AA8729161CE354777C4DB> (accessed on 01.06.2021)
7. The Well-being of Nations. The Role of Human and Social Capital. OECD, 2001. URL: <https://www.oecd-ilibrary.org/docserver/9789264189515-en.pdf?expires=1634023630&id=id&accname=guest&checksum=422F778380B5AB356A774DC93897A759> (accessed on 01.06.2021)
8. World Bank. The Human Capital Project: Frequently Asked Questions. URL: https://www.worldbank.org/en/publication/human-capital/brief/the-human-capital-project-frequently-asked-questions?cid=GGH_e_hcpexternal_en_ext#HCP2 (accessed on 15.06.2021)
9. Encyclopedia Britannica. Human Capital. URL: <https://www.britannica.com/topic/human-capital>. (accessed on 15.06.2021)
10. ISCO-88: International Standard Classification of Occupations. – Geneva, 1990.

11. International Standard Classification of Occupations: ISCO 08 / International Labour Office. -. Geneva: ILO, 2012. URL: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf (accessed on 15.06.2021)
12. ОК 010–2014 (MSKZ-08). Obshcherossiyskiy klassifikator zanyatiy (prinyat i vveden v deystviye Prikazom Rosstandarta ot 12.12.2014 N 2020-st). URL: <https://docs.cntd.ru/document/1200121893> (accessed on 01.06.2021)
13. Guide to Measuring Information and Communication Technologies (ICT) in Education. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (accessed on 01.06.2021)
14. International Standard Classification of Education: Fields of Education and Training 2013. Detailed field descriptions. UNESCO Institute for Statistics, Montreal, Quebec, 2015. URL: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-fields-of-education-and-training-2013-detailed-field-descriptions-2015-en.pdf> (accessed on 15.06.2021)
15. Indeks gotovnosti regionov Rossii k informatsionnomu obshchestvu. 2004–2005 / Pod red. T.V. Yershovoy, YU.Ye. Khokhlova, S.B. Shaposhnika. - M. : Institut razvitiya informatsionnogo obshchestva, 2005. - 224 s. URL: <http://eRegion.ru> (accessed on 01.06.2021)
16. Indikatory tsifrovoy ekonomiki: 2021: statisticheskiy sbornik / G. I. Abdrakhmanova, K. O. Vishnevskiy, L. M. Gokhberg i dr. ; Nats. issled. un-t «Vysshaya shkola ekonomiki». - M. : NIU VSHE, 2021. - 380 s.
17. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // Informatsionnoye obshchestvo. 2021. № 4–5, P. 2–32. https://doi.org/10.52605/16059921_2021_04_02
18. Measuring ICT usage and electronic commerce in enterprises: proposal for a model questionnaire. OECD, Paris, 11–12 October 2001. URL: <https://www.oecd.org/digital/ieconomy/20628443.pdf> (accessed on 01.06.2021)
19. ICT use by businesses: revised OECD model survey. OECD, Paris, 2005. URL: <https://www.oecd.org/digital/ieconomy/35867672.pdf> (accessed on 01.06.2021)
20. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (accessed on 01.06.2021)
21. Measuring the Information Society Report 2017. Volume 1 – Geneva: International Telecommunication Union, 2017. – xi + 155 p.
22. Extraordinary Meeting of the Expert Group on ICT Household Indicators (EGH) and Expert Group on Telecommunication/ICT Indicators (EGTI). URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/events/eghegti2017/default.aspx>. (accessed on 01.06.2021)
23. The ICT Development Index (IDI): Methodology, indicators and definitions (as of February 2019). International Telecommunication Union, 2019. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/ITU_IDI%20Development%20Index.pdf (accessed on 01.06.2021)
24. Measuring the Information Society Report 2018. Volume 1 – Geneva: International Telecommunication Union, 2018. – xi + 204 p.
25. Measuring the Information Society Report 2018. Volume 2 – Geneva: International Telecommunication Union, 2018. – viii + 244 p.
26. ICT Development Index 2020: A proposal. International Telecommunication Union, 2020. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/events/egti2020/IDI2020_BackgroundDocument_20200903.pdf (accessed on 15.06.2021)
27. United Nations E-Government Survey 2020. - UN, New York, 2020. - 323 p. URL: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020> (accessed on 15.06.2021)
28. Network Readiness Index 2020: Benchmarking the Future of the Network Economy. Portulance Institute, October 2020. – 317 p. URL: <https://networkreadinessindex.org/> (accessed on 15.06.2021)
29. The Digital Economy and Society Index (DESI). URL: <https://ec.europa.eu/digital-single-market/en/desi> (accessed on 15.06.2021)

30. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model. European Commission, 2016. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101254/jrc101254_digcomp%202.0%20the%20digital%20competence%20framework%20for%20citizens.%20update%20phase%201.pdf (accessed on 01.06.2021)
31. Individuals who have basic or above basic overall digital skills by sex. Eurostat Metadata. URL: https://ec.europa.eu/eurostat/cache/metadata/en/tepsr_sp410_esmsip2.htm. (accessed on 01.06.2021)
32. Natsional'nyy razvitiye tsifrovogo ekonomiki: Pilotnaya realizatsiya. Goskorporatsiya «Rosatom». Pod red. T.V. Yershovoy, YU.Ye. Khokhlova, S.B. Shaposhnika. 2018. - 92 s.
33. Artificial Intelligence Index Report 2021. Stanford University. Human-Centered Artificial Intelligence Institute (HAI). 2021. - 222 p.
34. Global AI Index. Tortoise Media. URL: <https://www.tortoisemedia.com/intelligence/global-ai/> (accessed on 01.06.2021)
35. Government AI Readiness Index 2020. Oxford Insights. 2020. - 144 p.
36. Righi, R., López-Cobo, M., Alaveras, G., Samoili, S., Cardona, M., Vázquez-Prada Baillet, M., Ziemba, L.W., and De Prato, G., Academic offer of advanced digital skills in 2019-20. International comparison. Focus on Artificial Intelligence, High Performance Computing, Cybersecurity and Data Science, EUR 30351 EN, Publications Office of the European Union, Luxembourg, 2020. URL: <https://op.europa.eu/en/publication-detail/-/publication/fad44166-f249-11ea-991b-01aa75ed71a1> (accessed on 01.06.2021)
37. Studyportal. URL: <https://studyportals.com/> (accessed on 15.06.2021)
38. Lea Samek, Mariagrazia Squicciarini and Emile Cammeraat The human capital behind AI: jobs and skills demand from online job postings. OECD, 2021. - 99 p. URL: https://www.oecd-ilibrary.org/science-and-technology/the-human-capital-behind-ai_2e278150-en (accessed on 15.06.2021)
39. *Burning Glass Technologies*. URL: <https://www.burning-glass.com/>. (accessed on 15.06.2021)
40. Sena, Vania & Ozdemir, Sena. (2020). Spillover effects of investment in big data analytics in B2B relationships: What is the role of human capital?. *Industrial Marketing Management*. 86. 77-89. <https://doi.org/10.1016/j.indmarman.2019.05.016>.
41. K. Rasheva-Yordanova, E. Iliev, V. Chantov, Analysis Of Missing Data Science Competence In It Sector” . Proceedings of EDULEARN18 Conference 2nd-4th July 2018, Palma, Mallorca, Spain IATED, ISBN: 978-84-09-02709-5, pp. 7399 – 7403, 201. <http://doi.org/10.21125/edulearn.2018.1731>
42. Jun Wu, Honglei Shi, and Jiaping Yang, “Are big data talents different from business intelligence expertise?: Evidence from text mining using job recruitment advertisements,” in 2017 International Conference on Service Systems and Service Management, Jun. 2017, pp. 1–6, ISSN:2161-1904. <https://doi.org/10.1109/ICSSSM.2017.7996289>
43. K. Rasheva-Yordanova, E. Iliev, B. Nikolova, “Analytical Thinking As A Key Competence For Overcoming The Data Science Divide,” in Proceedings of EDULEARN18 Conference 2-4 July 2018, Palma, Mallorca, Spain, pp 7892-7898, 2018. <http://doi.org/10.21125/edulearn.2018.1833>
44. Pilipczuk, O.; Cosenco, N.; Kosenko, O. Big Data: Challenges and opportunities in financial management. *Manag. Issues* 2019, 5, 9–23. <https://doi.org/10.7172/1644-9584.85.1>
45. Mottier, E., Marshall, T. An Investigation into the Importance of Analytical Competencies: a Hospitality Industry Perspective. In: Proceedings of The 11th EuroMed Conference, 12-14 September, 2018. ISBN: 978-9963-711-67-3.
46. Pilipczuk O. Toward Cognitive Management Accounting. *Sustainability* 2020, 12(12), 5108; <https://doi.org/10.3390/su12125108>
47. Hakimi, A.A.A., Ali, M.S.M., Wahid, H. Big Data Skills Required for Successful Application Implementation in the Banking Sector, in Modeling, Design and Simulation of Systems Communications in Computer and Information Science, Vol. 752, 381-392 - August, 2017. https://doi.org/10.1007/978-981-10-6502-6_34
48. Pesqueira, A., Sousa, M.J. & Rocha, Á. Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals. *J Med Syst* 44, 197 (2020). <https://doi.org/10.1007/s10916-020-01665-9>

49. A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for Big Data professions: A systematic classification of job roles and required skill sets," *Information Processing & Management*, vol. 54, no. 5, pp. 807–817, Sep. 2018. <https://doi.org/10.1016/j.ipm.2017.05.004>
50. Akhtar, P., Frynas, J.G., Mellahi, K. and Ullah, S. (2019), Big Data-Savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *Brit J Manage*, 30: 252-271. <https://doi.org/10.1111/1467-8551.12333>
51. K. Rasheva-Yordanova, "Forming of Data Science Competence for Bridging the Digital Divide," in *The Future Education*, 2018
52. Catherine Chen & Haoqiang Jiang (2020) Important Skills for Data Scientists in China: Two Delphi Studies, *Journal of Computer Information Systems*, 60:3, 287-296, <https://doi.org/10.1080/08874417.2018.1472047>
53. T. Wiktorski, Y. Demchenko, A. Belloum and A. Shirazi, "Quantitative and Qualitative Analysis of Current Data Science Programs from Perspective of Data Science Competence Groups and Framework," 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2016, pp. 633-638, <https://doi.org/10.1109/CloudCom.2016.0109>
54. J. C. Nwokeji, R. Stachel, T. Holmes, F. Aqlan, E. C. Udenze and R. Orji, "Panel: Addressing the Shortage of Big Data Skills with Inter-Disciplinary Big Data Curriculum," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-4, <https://doi.org/10.1109/FIE43999.2019.9028350>
55. Ukaz Prezidenta Rossiyskoy Federatsii ot 21 iyulya 2020 g. N 474 «O natsional'nykh tselyakh razvitiya Rossiyskoy Federatsii na do 2030 goda». URL: <http://publication.pravo.gov.ru/Document/View/0001202007210012> (accessed on 01.06.2021)

Измерение информационного общества

ИССЛЕДОВАНИЯ И РАЗРАБОТКИ В ОБЛАСТИ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Аннотация

Разработаны концептуальная схема и набор показателей для мониторинга исследований и разработок как фактора развития и использования технологий работы с большими данными. В концептуальной схеме мониторинга российских исследований и разработок в области больших данных выделены три основных компонента: масштабы, уровень, а также инновационные и международные связи. Произведен расчет показателей для России и сделан сравнительный анализ полученных результатов с другими странами.

Ключевые слова

большие данные, технологии работы с большими данными; исследования и разработки; публикационная активность; патентная активность; наукометрические показатели; мониторинг и оценка

Введение

Инновационность является характерной чертой цифровой экономика, поскольку она связана с развитием и внедрением современных цифровых технологий, которое сопровождается преобразованием деловых процессов, выводом на рынок новых товаров и услуг, изменением бизнес-моделей.

Опыт гигантов и стартапов цифровой индустрии говорит о том, что научный фундамент для современных инноваций в области информационно-коммуникационных технологий (ИКТ) становится обязательным и определяет успех развития компаний, даже если в период их создания интенсивные научные исследования не проводились. Характерно, что в базе данных Web of Science (WoS), в которой индексируются наиболее значимые научные публикации (прежде всего в области фундаментальных исследований), довольно быстро растет число публикаций сотрудников крупных цифровых компаний. Например, в первые годы существования компании Facebook это были единичные публикации, тогда как в последние годы (2017–2020) ежегодно в WoS индексируется несколько сотен публикаций (в т. ч. по технологиям работы с большими данными), подготовленных сотрудниками исследовательских подразделений. Исследования нужны компании Facebook, чтобы улучшать сервисы социальной сети, использовать таргетированную рекламу как один из основных способов монетизации, развивать новые направления бизнеса и т. д.

Наличие развитой сферы исследований и разработок в области технологий хранения и анализа больших данных необходимо также для обеспечения национальной безопасности – в ряде

© Хохлов Ю.Е. Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_90

чувствительных областей возможно использование только отечественных решений для работы с большими данными.

В данной статье представлены концептуальная схема и методология мониторинга исследований и разработок в области технологий работы с большими данными, а также результаты оценки этой области в 2020 году, проведенной на основе разработанной методологии (и части предложенных показателей).

1 Определение предметной области мониторинга и оценки

Предметом мониторинга и оценки в данном направлении выступает состояние исследований и разработок в области технологий работы с большими данными в Российской Федерации.

В документах стратегического планирования, которые приняты в последнее время (федеральный проект «Цифровые технологии», дорожные карты по развитию сквозных цифровых технологий), заложены традиционные представления об инновационной цепочке: результаты отечественных научных исследований должны использоваться при разработке технологических решений, которые, в свою очередь, коммерциализируются, выводятся на российский и зарубежные рынки и внедряются в различные сферы деятельности, приводя к их трансформации. Хотя эта логика не покрывает всех механизмов инноваций – научные исследования не являются их единственным источником, – в ней представлена значительная часть инноваций, наиболее существенных и прорывных. Поэтому данную предметную область мониторинга можно определить как исследования и разработки в области технологий работы с большими данными, находящиеся на начальных уровнях готовности цифровых технологий (УГТ 1–4 в терминах шкалы уровней готовности технологий, см. [1, 2]). Результаты исследований и разработок в этой области как правило представлены в виде научных публикаций и патентов, измеримые характеристики которых используются при мониторинге и оценке.

Для операционализации определения предметной области «Исследования и разработки в области технологий работы с большими данными» необходимо охарактеризовать ее с помощью действующих классификаторов. Сразу отметим, что анализ подобных (достаточно узких направлений) исследований и разработок затруднен тем, что существующие классификаторы (научных областей, патентов) достаточно «грубы» и не предоставляют возможности выделить предметную область путем формального перечисления кодов существующих классификаторов. В данном случае для мониторинга исследований в области технологий работы с большими данными, основанного на анализе массивов публикаций и патентов был реализован подход, включающий два следующих этапа.

1. Определение более широкой, формально определяемой предметной области исследований и разработок, в рамках которой будет выделяться более узкая область, связанная с технологиями работы с большими данными. Такой широкой предметной областью для целей настоящей работы являются исследования и разработки в области ИКТ.

Организация экономического сотрудничества и развития (ОЭСР) разработала классификатор областей науки и технологий (Fields of Science and Technology, FOS 2007) [3], с которым в настоящее время гармонируются отечественные классификаторы. В рамках данной работы для определения массива публикаций в области ИКТ использованы две области FOS 2007 на 2-м уровне классификатора: “1.2 Computer and information sciences” и “2.2 Electrical engineering, Electronic engineering, Information engineering”.

Для выделения массива патентов в области ИКТ предлагается использовать Международную патентную классификацию и группировку ее кодов, относящихся к ИКТ, которая предложена ОЭСР для мониторинга развития информационного общества [4].

2. Для выделения узкой предметной области, связанной с исследованиями и разработками по технологиям работы с большими данными, сформирован перечень ключевых слов и поисковых образов, позволяющих выбрать в международных базах данных публикаций и патентов релевантные документы (см. раздел 4).

Массив публикаций и патентов, полученный на основе сформированных поисковых образов, анализируется в соответствии с набором показателей из концептуальной схемы мониторинга (см. раздел 2).

2 Концептуальная схема и показатели мониторинга исследований и разработок в области работы с большими данными

2.1 Обзор международных стандартов и показателей мониторинга сферы исследований и разработок

Международные стандарты статистического наблюдения (мониторинга) сферы исследований и разработок были предложены и начали использоваться еще в 1960-х годах. В 1963 г. ОЭСР впервые опубликовала методологию сбора статистических данных об исследованиях и разработках – Руководство Фраскати [5]. Это руководство обязано своим названием итальянскому городу, где в 1962 году Рабочая группа национальных экспертов ОЭСР по показателям науки и техники впервые согласовала общий подход к измерению и представлению статистических данных о НИОКР. Руководство пересматривалось семь раз для решения новых задач и учета возникающих интересов пользователей, 7-е издание опубликовано в 2015 г. Руководство Фраскати в своей статистической практике используют не только страны ОЭСР, но и большинство стран мира; российское статистическое наблюдение за сферой исследований и разработок также гармонизировано со стандартами ОЭСР. В руководстве даны определение и указания по измерению таких аспектов науки, как типы исследований и разработок (фундаментальные исследования, прикладные исследования, разработки), научный персонал (исследователи, техники, вспомогательный персонал), области науки и сектора проведения научных исследований (бизнес, вузы, государство, частные некоммерческие организации). Показатели, основанные на Руководстве Фраскати, получили впоследствии название показатели «входа», так как характеризуют кадровые, финансовые и организационные ресурсы, выделяемые на проведение исследований и разработок.

В 1960-е годы начал активно развиваться еще один подход к мониторингу и исследованию научной деятельности: разработка и использование библиометрических показателей, характеризующих научный «выход» – результаты исследований и разработок, представленные в научных публикациях и патентах. Важным стимулом развития этого направления стало создание в 1961 г. Института научной информации (ISI, США), который вышел на рынок научной информации с библиографической системой «Индекс научного цитирования» (Science Citation Index). В настоящее время она принадлежит компании Clarivate Analytics и известна под торговой маркой Web of Science [6].

Возможность анализа науки с точки зрения «выхода» и значимость для развития этого направления индексов научного цитирования основана на специфических особенностях научной деятельности. В этой связи достаточно упомянуть следующие обстоятельства.

1. Основными «продуктами» научной деятельности являются научные публикации и патенты, в которых фиксируются и обосновываются результаты и методы исследований и разработок. Достоверное отражение полученных результатов научной деятельности в информационной системе науки является социальной нормой, соблюдение которой достаточно строго контролируется.

2. Большие возможности для реализации подхода дает отражение в библиографических системах научных ссылок индексируемых публикаций. В любой научной работе используются результаты предыдущих исследований: каждый научный результат влетает в единую ткань развивающегося научного знания. При подготовке публикации научный работник обязательно помещает свои результаты в контекст развития определенной исследовательской области, ссылаясь на наиболее существенные для него научные публикации. Адекватное отражение в библиографических ссылках использованного в работе познавательного аппарата и результатов предыдущих исследований регулируется довольно жесткими нормами. Поэтому «систему библиографических ссылок» можно рассматривать как особую, очень информативную семиотическую систему, пригодную для решения целого ряда задач, в том числе для оценки состояния исследований, определения структуры и приоритетов переднего фронта исследований, определения публикаций и журналов, которые оказывают наибольшее воздействие на развитие науки, анализа инновационных связей и т. д.

Показатели «входа» и «выхода» широко используются в настоящее время в национальных и международных системах мониторинга, а также в международных рейтингах развития сферы научно-исследовательских и опытно-конструкторских работ (НИОКР). Эти показатели можно объединить в группы для характеристики:

- масштабов национальных научных систем (доля затрат на исследования и разработки в ВВП страны, доля исследователей в населении, количественные и нормированные показатели публикационной и патентной активности);
- секторальной и дисциплинарной структуры затрат и научного персонала (доля в расходах на НИОКР государства, бизнеса, некоммерческого сектора в т.ч. в разрезе научных дисциплин и направлений);
- уровня исследований (цитируемость публикаций – средняя цитируемость, доля высокоцитируемых статей, публикации в ведущих рецензируемых журналах из квартилей Q1 и Q2 и др.);
- международного сотрудничества (доля публикаций в международном соавторстве);
- инновационных связей (цитирование научных публикаций в патентах, сотрудничество организаций академического сектора и бизнеса).

В контексте данной работы важно отметить, что в системах мониторинга и оценки технологического развития показатели состояния НИОКР используются как характеристика одного из факторов развития. Это относится как к оценкам общего технологического и инновационного развития (например, в Глобальном индексе инноваций [7]), так и определенного направления технологического развития (например, в индексах развития и использования технологий искусственного интеллекта [8]). Публикационная активность в них измеряется, как правило, на основе библиографических баз данных Web of Science и Scopus, патентная активность – на основе баз данных патентов, а для международных сравнений – на основе наиболее полных международных систем патентной информации (например, Patentscope Всемирной организации интеллектуальной собственности [9]).

Если речь идет о более узких технологических направлениях, таких как развитие и использование технологий работы с большими данными, определенную проблему составляет его выделение в статистике «входа» и «выхода» – действующие классификаторы, как правило, достаточно грубы и не позволяют охарактеризовать его напрямую. Несколько подходов к решению этой проблемы для показателей «выхода» рассмотрены в следующих разделах.

2.2 Концептуальная схема предметной области

Результаты анализа из предыдущих разделов позволяют задать концептуальную схему мониторинга исследований и разработок в области технологий работы с большими данными в виде трех основных компонентов: (1) масштабы исследований и разработок; (2) уровень национальных исследований по сравнению с мировым; (3) международные и инновационные связи, трансфер технологий (см. таблицу 1).

Таблица 1 - Концептуальная схема мониторинга исследований и разработок в области технологий работы с большими данными

Масштабы исследований и разработок	Уровень исследований и разработок	Международные и инновационные связи
Публикационная активность	Цитируемость работ	Сотрудничество научных организаций и бизнеса
Патентная активность	Публикации в ведущих журналах	Использование результатов исследований в технических изобретениях
Количество исследователей		Международное сотрудничество

Масштабы исследований и разработок традиционно измеряются показателями вооруженности экономики знаниями и техническими изобретениями (массивы публикаций и патентов страны, отнесенные к размеру национальной экономики) и ее обеспеченности исследовательскими кадрами и финансовыми ресурсами, выделяемыми на НИОКР (доля расходов

на НИОКР в ВВП, доля исследователей в занятом населении) [7]. В данной работе речь будет идти о масштабах исследований и разработок в области технологий работы с большими данными.

Помимо масштаба исследований и разработок важно оценить их «качество». Уровень научных исследований характеризуется общепринятыми (в том числе принятыми органами управления наукой) показателями, связанными с цитируемостью работ и публикациями в ведущих рецензируемых журналах.

Для развития цифровой экономики важно использование фундаментальных знаний для разработки технологий и решений, поэтому для мониторинга важно оценить инновационные связи и трансфер технологий, то есть понять, насколько проводимые исследования используются для технических изобретений, а разработки – бизнесом. Этот компонент предметной области мониторинга в свою очередь содержит два направления оценки: (1) научное сотрудничество научно-образовательных организаций (вузов и НИИ) с бизнесом, (2) использование результатов исследований в технических изобретениях. Важным аспектом в этом компоненте предметной области служит характеристика международных научных связей, которые являются каналом обмена знаниями, исследовательскими практиками и техническими идеями.

2.3 Показатели масштабов исследований и разработок

Научные публикации и патенты являются основными результатами научных исследований и разработок, поэтому измерение их количественных характеристик регулярно используется при мониторинге – от оценки объемов проводимых исследований до анализа структуры переднего фронта исследований и оценки результатов научной деятельности. Для целей данной работы в мониторинг масштабов исследований в области технологий работы с большими данными включены библиометрические показатели публикаций российских ученых, индексированных в WoS, и российских патентных заявок, поданных в соответствии с договором о патентной кооперации (Patent Cooperation Treaty, PCT) и опубликованных в международных базах данных.

Патенты – это право интеллектуальной собственности, выданное уполномоченными органами изобретателям, позволяющее им использовать свои изобретения и эксплуатировать их в течение ограниченного периода времени (обычно 20 лет). Патенты выдаются частным лицам, фирмам или другим юридическим лицам при условии, что изобретение соответствует определенным критериям: оно должно быть новым, включать изобретательский уровень (то есть быть неочевидным) и быть пригодным для промышленного применения.

Патенты являются одним из ключевых показателей инновационной деятельности. Они используются для измерения результатов НИОКР, распространения знаний, изобретательских достижений в качестве инструмента для оценки направления исследований и стратегических целей компаний и стран. Анализ массива патентов может также дать представление о связях технических изобретений и научных исследований.

Для характеристики масштабов исследований в области технологий работы с большими данными предлагаются показатели, аналогичные включаемым в международные рейтинги и системы мониторинга состояния научно-инновационной сферы.

1. Прежде всего, это «Доля российских публикаций в области технологий работы с большими данными, %» в системе WoS.
2. Для корректных международных сопоставлений публикационная активность относится к размеру национальной экономики – используется удельный показатель «Количество публикаций по технологиям работы с большими данными на 1 млрд ВВП». Этот показатель можно трактовать как характеристику обеспеченности экономики исследованиями в области технологий работы с большими данными.
3. Для характеристики масштабов разработок в области технологий работы с большими данными и проведения международных сопоставлений, как и в случае с публикациями, используется показатель «Количество патентных заявок по технологиям работы с большими данными, поданных по процедуре PCT, на 10 млрд ВВП».

Наукоемкость цифровой экономики характеризует еще один показатель, связанный с обеспеченностью исследований и разработок (в данном случае в области технологий работы с большими данными) кадровыми ресурсами. Учитывая, что в классификаторах областей науки, используемых для статистических наблюдений за кадровым составом науки, отсутствует специализация в области больших данных, для измерения этого важного аспекта сферы

исследований и разработок применен подход, основанный на анализе публикационной активности в этой области. В рамках этого подхода в качестве прокси-метрики для измерения обеспеченности исследователями в области больших данных предлагается использовать следующий показатель:

4. Количество уникальных авторов публикаций по технологиям работы с большими данными в WoS на 100 тыс. занятых.

2.4 Показатели уровня исследований и разработок

При мониторинге и оценке состояния научных исследований важно иметь не только объемные показатели, но и уровень исследований. Для этого предлагается использовать следующие показатели цитируемости публикаций.

1. «Средняя цитируемость публикаций, нормализованная по области исследований, году и типу публикаций». Этот показатель рассчитывается как среднее значение нормализованных показателей цитируемости публикаций страны, организации и т. п. Нормализация осуществляется делением числа цитирований, полученных публикацией, на среднюю по миру цитируемость работ в данной области того же года и типа. Такой подход позволяет при оценке убрать зависимость числа цитирований от года выхода публикации (более ранние в среднем цитируются выше – набирают больше ссылок), типа (например, обзоры цитируются чаще) и области исследований (в более «крупных» областях средняя цитируемость выше). При интерпретации значений данного показателя исходят из следующего: если показатель меньше 1, то цитируемость ниже среднемировой, если больше – выше. Тем самым показатель позволяет оценить, насколько востребованность результатов исследований страны (или организации) отличается от среднемировой.
2. «Доля высокоцитируемых публикаций среди всех публикаций по технологиям работы с большими данными». Присутствие высокоцитируемых публикаций свидетельствует о наличии прорывных исследований, оказывающих большое воздействие на отечественную и мировую науку.

2.5 Показатели международных и инновационных связей

Возможности научных исследований генерировать новые технологии характеризуют наличие устойчивых связей научных исследований и изобретательской активности, связей научных организаций с предпринимательским сектором экономики, а также международное сотрудничество, которое является важным каналом обмена знаниями, исследовательскими практиками и возможностями. Для мониторинга и оценки этого компонента предметной области введены следующие показатели.

1. «Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией». Показатель характеризует трансфер технологий из академического сектора в коммерческий. Он является одной из проблемных зон российской инновационной системы – в целом по компьютерным наукам доля российских научных публикаций, выполненных в соавторстве с индустрией, на порядок меньше, чем у стран с эффективными инновационными системами, высокой инновационной активностью предприятий и успешной коммерциализацией разработок (Швеция, Швейцария, Израиль, США) [23].
2. «Доля публикаций, на которые ссылаются патенты в области технологий работы с большими данными». Данный показатель отражает способность научных исследований служить основой для разработки новых технологий.
3. «Доля публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве». Участие в международном научном сотрудничестве обогащает потенциал национальной науки, позволяет обмениваться исследовательскими практиками, которые в значительной мере являются «неявным знанием» и не представлены в научных публикациях. Сотрудничество с ведущими мировыми исследовательскими центрами является и косвенным свидетельством уровня исследований. Совместная научная деятельность обычно проявляется в совместных публикациях, поэтому использование показателя публикаций с международным соавторством является наиболее естественной операционализацией понятия международного научного сотрудничества.

3 Методология измерения показателей

3.1 Обзор библиометрических исследований, посвященных анализу публикационной и патентной активности в области технологий работы с большими данными

Как уже отмечалось, проблемой для измерения показателей исследований и разработок в области технологий работы с большими данными является выделение этой области с точки зрения имеющихся классификаторов. Действующие классификаторы (областей науки, патентов, публикаций в библиографических базах данных) достаточно «грубы», не отражают эту узкую область и не позволяют измерять описанный выше спектр показателей на основе действующего статистического наблюдения (показатели «входа») и стандартных поисковых инструментов баз данных (показатели «выхода»).

В этой ситуации одним из основных подходов является опора на показатели «выхода», для которых указанная проблема разрешима за счет использования ключевых слов, характеризующих данную область. С их использованием возможно выделение массива публикаций и патентов, которые относятся к технологиям работы с большими данными, что позволяет измерить не только библиометрические показатели «выхода», но и в некоторых случаях «ресурсные» показатели. Примером вычисления показателей «входа» на основе библиометрических показателей является подсчет уникальных авторов публикаций в анализируемой узкой области, что может выступать заменой показателя числа исследователей в этой области.

Такой подход к измерению исследований и разработок в области технологий работы с большими данными (и в других аналогичных узких областях), основанный на выделении массива публикаций с использованием ключевых слов и дальнейшем изучении его библиометрических характеристик, достаточно распространен.

За последние 20 лет (2001–2020) в базе данных WoS было проиндексировано 754 публикации, в названиях, ключевых словах и аннотациях которых встречаются одновременно термины, характеризующие технологии работы с большими данными (см. ниже раздел «Методология») и термин «bibliometrics» в различных вариантах написания. Основной массив публикаций (676) появился в последние 10 лет, когда, собственно, и начались активные исследования в области работы с большими данными. При этом половина публикаций последних 10 лет приходится на 2019 и 2020 гг. (144 и 193 публикации соответственно). Отметим, что при этом в большинстве указанных публикаций библиометрические инструменты использованы для обзора литературы при изучении отдельных аспектов работы с большими данными.

Для выделения собственно библиометрических работ, посвященных измерению исследований и разработок в области технологий работы с большими данными, параметры поиска были уточнены по следующим направлениям.

1. Поиск проводился только по названиям и ключевым словам публикаций последних 10 лет.

2. Для поиска использовался только термин «big data», а не весь набор специализированных терминов, так как интерес представляла методика выделения данной области исследований (279 публикаций за последние 10 лет). Дополнительно проводился поиск публикаций, у которых в названиях и ключевых словах одновременно используются термины “big data” и “patent” – для выявления статей, анализирующих патентную активность в области технологий работы с большими данными, но не использующих термин «bibliometrics» (46 публикаций за последнее десятилетие).

3. Проводился поиск только научных статей, содержащих развернутое описание методологии библиометрических исследований. Обзорные публикации не рассматривались, так как чаще всего они посвящены отдельным аспектам работы с большими данными в различных сферах деятельности.

4. В результате уточнения поискового запроса по приведенным выше критериям были найдены 58 публикаций, а после ознакомления с полными текстами были отобраны 12 релевантных публикаций [10–21], посвященных библиометрическим исследованиям публикационной и патентной активности в области больших данных. Анализ 46 публикаций, в которых используются термины «big data» и «patent», позволил отобрать еще одну публикацию, посвященную анализу патентной активности [22]; еще одна статья входила в ранее отобранные 12 релевантных публикаций; большинство же из указанных 46 публикаций посвящены использованию методов работы с большими данными для анализа патентной информации.

Анализ отобранных публикаций с точки зрения интересующих нас поисковых стратегий показал следующее.

1. Для выделения с помощью ключевых слов исследований и разработок в области технологий работы с большими данными используется, как правило, только ключевое слово «big data» (см., например [10, 11]), в одном случае – с добавлением синонима «mega data» [12]. В то же время, в статье Лью и соавторов [13] при формировании набора ключевых слов для поискового образа проводилось специальное исследование, включающее предварительный анализ релевантных публикаций и формирование исходного перечня специфичных для данной области ключевых слов с участием экспертов, а затем проведение широкого опроса экспертов с целью верифицировать и уточнить список. В результате был сформирован следующий перечень ключевых слов для построения поискового образа: big data, data analytics, data center, data science, data warehouse, Hadoop, large dataset, machine learning, MapReduce, NoSQL, predictive analytics, sentiment analysis, unstructured data.

2. Для анализа публикационной активности в области больших данных используются, как правило, базы данных WoS и Scopus. Первая отмечается как содержащая тщательно отобранные более качественные источники, вторая – как более широкая по охвату публикаций [14]. Для исследований, содержащих показатели цитирования, чаще используется база данных WoS. Для анализа патентной активности используются базы данных с наибольшим охватом патентных документов (например, Lens Open Source Platform [20]).

3. Поиск по ключевым словам проводится в различных поисковых полях. Используются три основных подхода: (а) поиск по теме исследования (одновременно в названиях, ключевых словах, аннотациях [11]); (б) поиск в названиях и ключевых словах [15]; (в) поиск только в названиях публикаций [12]. Наиболее часто проводится поиск в названиях и ключевых словах, а также по теме исследования. В патентных базах данных обычно используется поиск по текстовым полям – например, одновременно в названиях, рефератах и формуле изобретения [20].

4. При выделении полного массива публикаций важным является выбор типа публикаций для анализа. В отмеченных статьях [10–21] присутствует несколько подходов: (а) анализ только научных статей [13]; (б) анализ нескольких типов публикаций (например, “article”, “proceedings paper”, “editorial material”, “review”, “meeting abstract” [16]). Наиболее часто для анализа публикационной активности выбирают статьи, особенно при изучении совместных исследований, связей областей исследования или тематических трендов (см. также [17]).

3.2 Методология измерения показателей мониторинга исследований и разработок в области работы с большими данными

Выбор WoS для измерения и мониторинга публикационной активности связан с тем, что эта система является наиболее авторитетной международной библиографической базой, для индексации в которой тщательно отбираются рецензируемые журналы и другие научные издания. Показатели, основанные на анализе публикаций, индексированных в WoS, используются при оценке результатов научной деятельности и проведении широкого круга науковедческих исследований. Важным является то обстоятельство, что для анализа публикаций, представленных в WoS, разработаны аналитические инструменты, позволяющие получать широкий спектр библиометрических показателей (InCites). Наряду с этим, для целей данной работы также использовались библиографическая база данных Scopus и соответствующий аналитический инструментарий SciVal. База данных Scopus является конкурентом WoS, близка ей по функциональным возможностям и целям разработки, при этом в ней индексируется более широкий спектр источников. Выбор базы данных Patentscope Всемирной организации интеллектуальной собственности (ВОИС) для измерения показателей патентной активности определялся тем, что ВОИС является официальным публикатором международных патентных заявок, поданных по процедуре PCT, а используемые показатели основаны на их подсчете.

Ключевым вопросом для наукометрического анализа исследований в области технологий работы с большими данными является выделение публикаций, относящихся к данной предметной области, из всего публикационного потока, индексированного в используемых библиографических базах данных. Как отмечалось ранее при обзоре проведенных исследований других авторов, обычным подходом является использование ключевого слова “big data” для отбора публикаций, имеющих это словосочетание в названиях или ключевых словах. Подобный подход существенно сужает выборку и приводит к потере значительного количества релевантных публикаций.

Отказ в данной работе от использовавшегося ранее подхода связан как с общими соображениями, так и с результатами проведенных экспериментов и экспертных процедур. Общая гипотеза, подтвержденная в ходе поисковых экспериментов, состоит в том, что в достаточно развитых и обширных областях исследования публикации, посвященные специальной узкой тематике (или технологии), зачастую перестают упоминать в названиях и ключевых словах более широкую «материнскую» область исследований, но характеризуют тематику исследования специфическими для нее терминами. Тем самым, работа, оставаясь в рамках обширной предметной области исследований (в данном случае – больших данных), теряет в метаописаниях (ключевых словах, аннотациях, названиях) прямую терминологическую связь с ней и характеризуется более специальным набором терминов. Одним из примеров в нашем случае является такая специфическая технология работы с большими данными, как Hadoop, – из 1794 публикаций, проиндексированных за последние 5 лет в WoS и содержащих в названиях или ключевых словах термин “Hadoop”, более половины (1022) не содержат одновременно термина “big data”.

С учетом сказанного, для формирования массива публикаций в области технологий работы с большими данными был реализован следующий подход:

1. На первом этапе на основе консультаций с экспертами в области технологий работы с большими данными и анализа публикаций был сформирован исходный список ключевых слов, характеризующих данную область. Основными критериями отбора терминов для ключевых слов были:
 - a. релевантность – термин используется в рамках области технологий работы с большими данными;
 - b. полезность для целей данного исследования – термин расширяет искомый массив и позволяет идентифицировать публикации, которые не имеют в названии или других метаданных термина “big data”;
 - c. отсутствие «шума» в результатах поиска – т.е. основная часть публикаций, содержащих предлагаемый термин в названии, ключевых словах или аннотации, должна относиться к указанной области исследований и разработок.
2. На втором этапе предложенный список был верифицирован в ходе опроса экспертов по работе с большими данными. Предложенный набор ключевых слов был подтвержден, наибольшее число баллов, помимо “big data”, набрали ключевые слова “Hadoop” и “MapReduce”.
3. На основе разработанного списка ключевых слов был сформирован запрос для поиска в библиографических базах данных, учитывающий различные варианты написания терминов.

Дополнительно было принято решение использовать для поиска не только названия публикаций и ключевые слова, но и аннотации. Проведенные эксперименты и экспертная оценка поисковых выдач показали, что подавляющее число публикаций, имеющих в аннотациях искомые ключевые слова, относятся к области технологий работы с большими данными.

В итоге был сформирован следующий поисковый образ для публикаций в области технологий работы с большими данными:

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Указанный поисковый образ использовался для формирования массива публикаций за 2016–2020 годы. Поиск проводился в поле «Тема» (TS – поиск по названиям, ключевым словам и аннотациям) в базе данных WoS, в аналогичном поле в базе данных Scopus и в текстовых полях патентной базы Patentscope.

Библиометрический анализ исследований и разработок в области технологий работы с большими данными был проведен на массиве публикаций, выделенных по описанной выше методике и относящихся только к двум областям науки, которые связаны с ИКТ (по классификатору FOS 2007): “1.2 Computer and information sciences” и “2.2 Electrical engineering, Electronic engineering, Information engineering”.

4 Результаты

Описанная выше методология мониторинга исследований и разработок в области технологий работы с большими данными была протестирована в ходе пилотной реализации, в рамках которой было измерено большинство показателей «выхода», основанных на анализе публикационной активности. Измерение проводилось в основном на массиве публикаций за 2019 год, т.к. индексирование работ, опубликованных в 2020 году, на момент подготовки публикаций не было еще завершено. Анализ полученных результатов представлен в данном разделе.

4.1 Масштабы исследований и разработок

В 2019 году Россия занимала 17-е место среди стран по общему числу публикаций в области технологий работы с большими данными. Доля России от общего числа публикаций – 1,53%, что в двадцать раз меньше, чем у стран-лидеров (Китая и США – см. рис. 1). При этом Россия отстает по этому показателю от таких стран, как Бразилия, Пакистан и Саудовская Аравия.

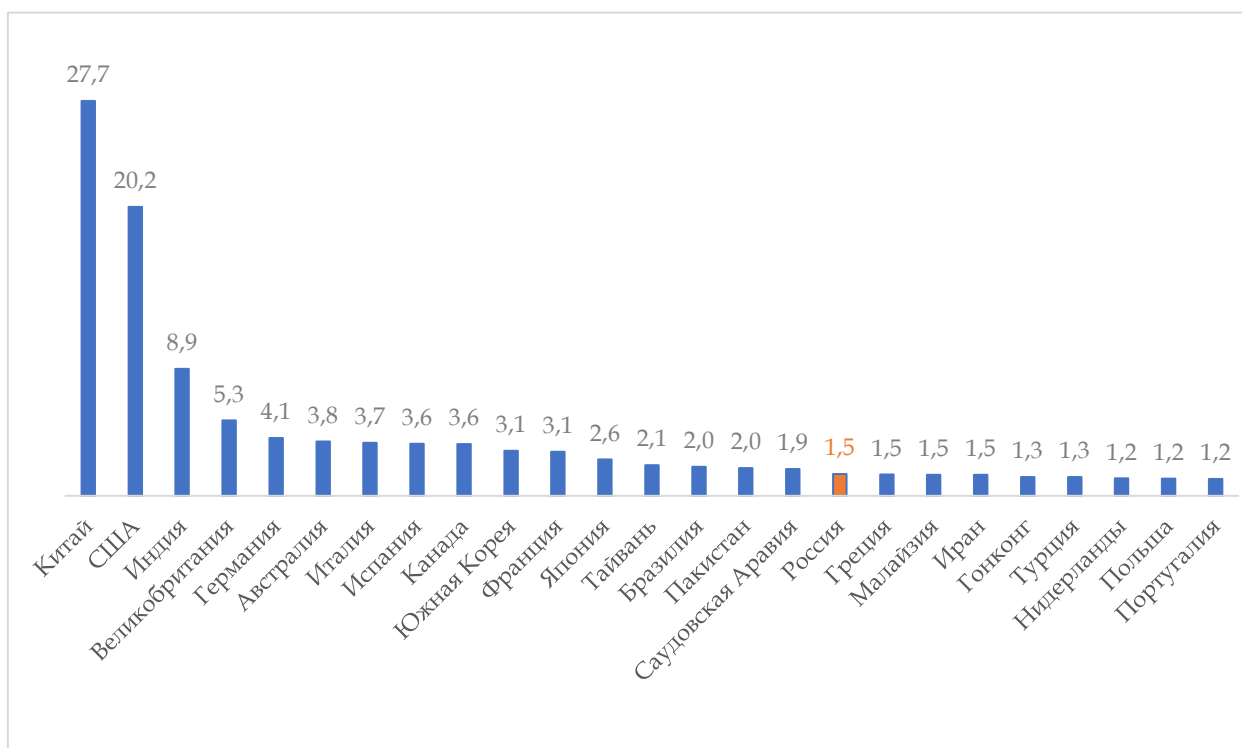


Рисунок 1. Доля публикаций России и стран мира от общего числа публикаций в области технологий работы с большими данными, индексируемых в WoS (25 стран-лидеров по числу публикаций, 2019 г.), %

Если посмотреть на динамику (рис. 2), то доля российских работ по сравнению с 2016 г. немного выросла (была 1,32%). В целом с точки зрения числа публикаций в области технологий работы с большими данными у пятерки стран-лидеров динамика разнонаправленная. У Китая продолжается устойчивый рост доли публикаций (более чем на 3 процентных пункта с 2016 г. по 2019 г.); США несколько снизили свое присутствие (на 1 процентный пункт с 2016 г. по 2019 г.); падает доля Индии; отмечается небольшой рост у Великобритании; доля Германии сначала уменьшилась, но последние 3 года медленно, но стабильно, растет. По неполным данным в 2020 г. эти тенденции в целом сохраняются и даже усиливаются – Китай существенно нарастил долю своих публикаций на фоне явно выраженного падения данного показателя у США.

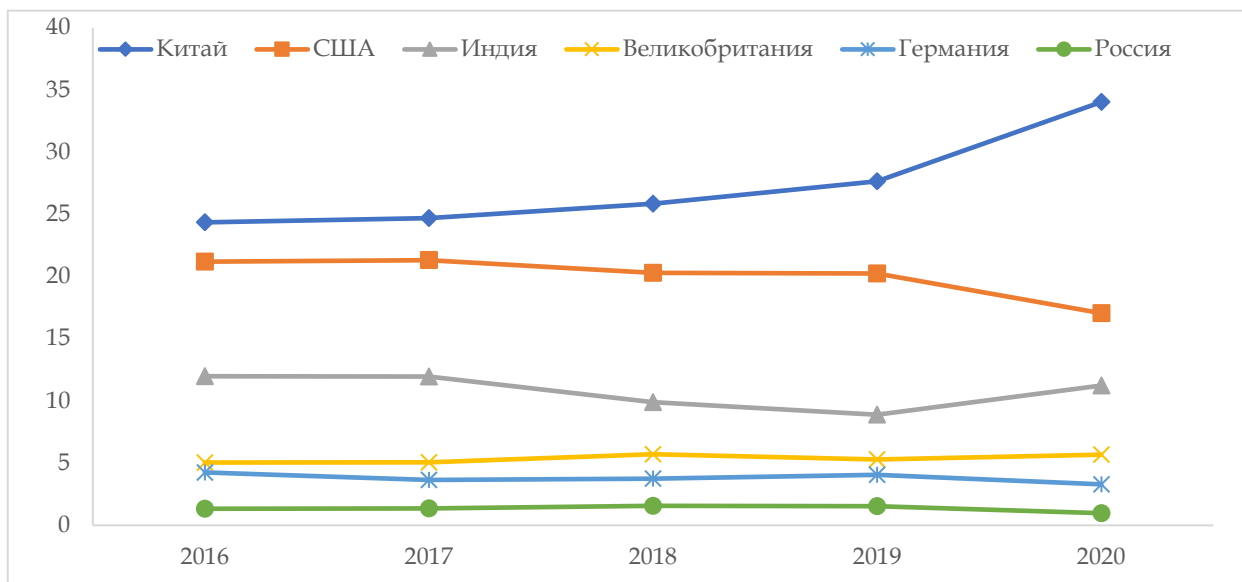


Рисунок 2. Динамика доли публикаций России и 5 стран-лидеров по числу публикаций в области технологий работы с большими данными, индексируемых в WoS, 2016–2020 гг., %

Для более корректных международных сопоставлений, учитывающих масштабы экономики сравниваемых стран, при мониторинге используется удельный показатель «Количество публикаций по технологиям работы с большими данными на 1 млрд ВВП». На рисунке 3 представлены рассчитанные для 2019 г. значения этого показателя для 25 стран с наибольшим числом публикаций в области технологий работы с большими данными. Для расчета использовались данные МВФ о размерах национальных экономик, рассчитанных по паритету покупательной способности (ППС).

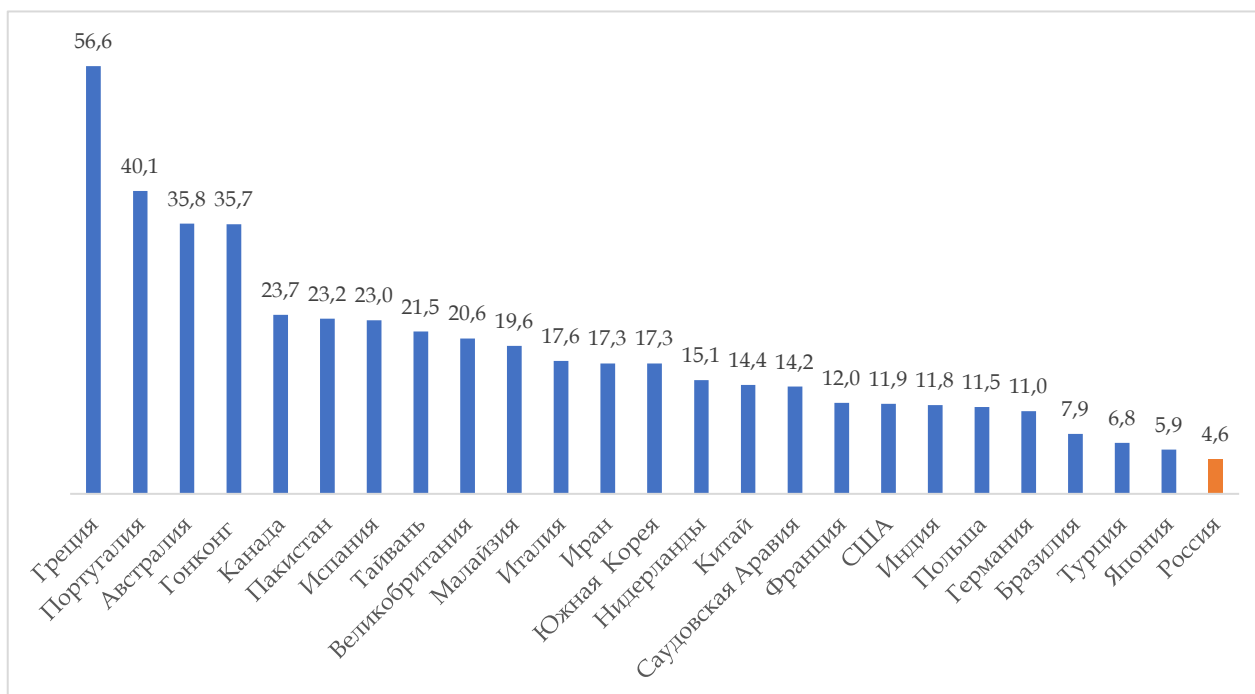


Рисунок 3. Количество публикаций в области технологий работы с большими данными на 1 млрд. ВВП по ППС, индексируемых в WoS (25 стран-лидеров по числу публикаций, 2019 г.), шт.

Россия в 2019 г. на порядок отставала от стран-лидеров по этому показателю и существенно уступала даже странам с большими экономиками (США, Китай, Индия, Германия).

Для мониторинга патентной активности были проанализированы сведения по международным патентным заявкам, поданным по процедуре РСТ. Поиск проводился в патентной базе данных Patentscope ВОИС. Учитывая, что заявки по процедуре РСТ могут подаваться на национальных языках участников договора (включая русский), для корректных международных сопоставлений поиск российских заявок осуществлялся на русском и английском языках, а для сравнения выбирались англоязычные страны. Поиск проводился по ключевым словам во всех текстовых частях патентных заявок. Для нормализации показателей использовались данные МВФ о размерах ВВП в международных долларах по паритету покупательной способности [24].

На рисунке 4 приведены рассчитанные за 2016–2020 гг. значения показателя «Количество патентных заявок, поданных по процедуре РСТ, по технологиям работы с большими данными на 10 млрд ВВП» для России и англоязычных стран.

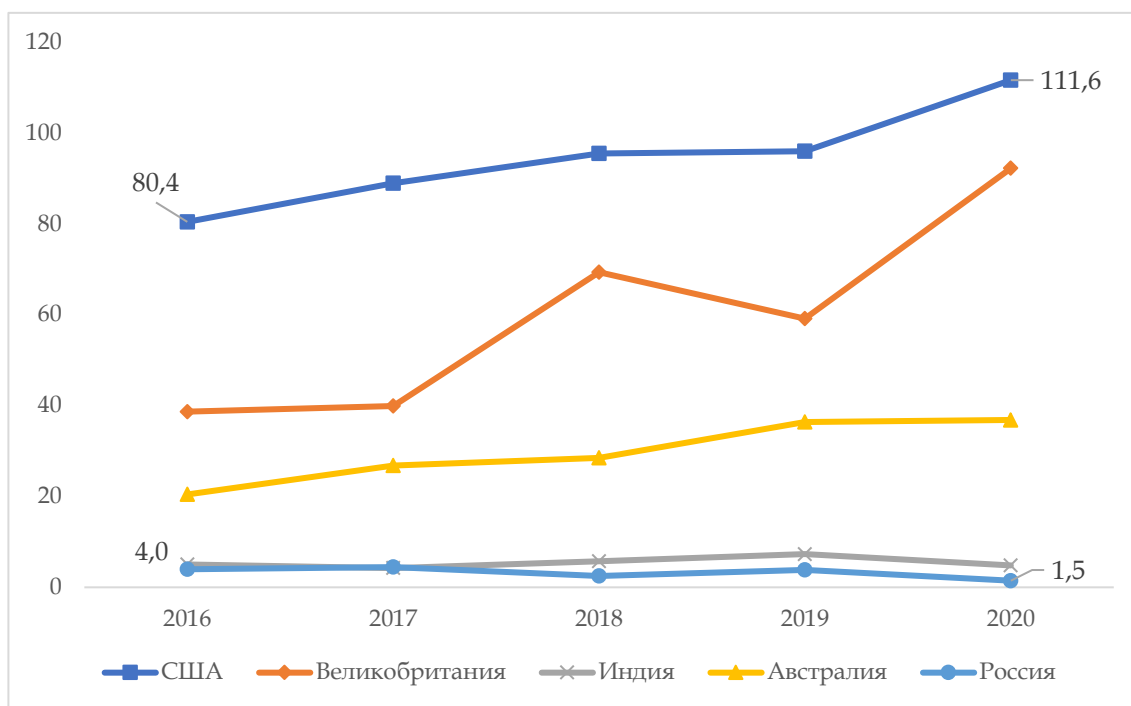


Рисунок 4. Количество международных патентных заявок в области технологий работы с большими данными, поданных по процедуре РСТ, на 10 млрд долларов ВВП по ППС (Россия и англоязычные страны, 2016–2020 гг.), шт.

Россия в 2020 г. почти на два порядка отставала от США по этому показателю (в 76 раз), в остальные годы разрыв варьировался от 20 до 38 раз. Значения показателя у России во все анализируемые годы были близки к индийским и, в отличие от других сравниваемых стран, не имели тенденции к росту – значения показателя у США за рассматриваемый период выросли почти в 1,5 раза, а у Великобритании – более чем в 2 раза.

Как и в случае публикационной активности, в последние годы в области подачи патентных заявок по технологиям работы с большими данными доминирует Китай. На рисунке 5 приведены данные за 2020 г. по доле патентных заявок, поданных по процедуре РСТ, у которых в названии или реферате заявки встречаются разработанные для данного исследования ключевые слова (см. выше раздел 4). Ограничение поиска только по названиям и рефератам в данном случае вызвано необходимостью проведения международных сопоставлений – во всех заявках РСТ, даже поданных на национальных языках, есть название и реферат на английском языке (отметим еще раз, что приведенные выше на рисунке 4 данные – результат поиска по всем текстовым полям заявок, поэтому данные по странам различаются).

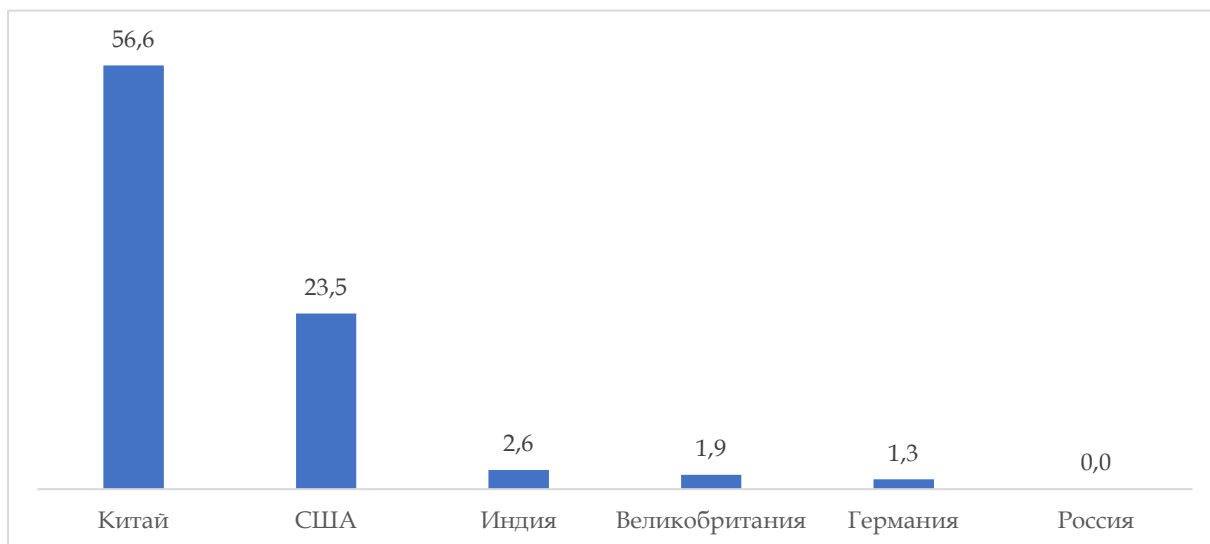


Рисунок 5. Доля международных патентных заявок, поданных по процедуре РСТ, в названии и реферате которых встречаются ключевые слова технологий работы с большими данными (2020 г.), %

5.2 Уровень исследований

Среди 25 стран с наибольшим числом публикаций в области технологий работы с большими данными Россия в 2019 г. занимает 16-е место по средней цитируемости публикаций, нормализованной по области исследования, году и типу публикаций. Ее показатель равен 1.21, а это означает, что цитируемость российских работ выше среднемировой (см. рис. 6).

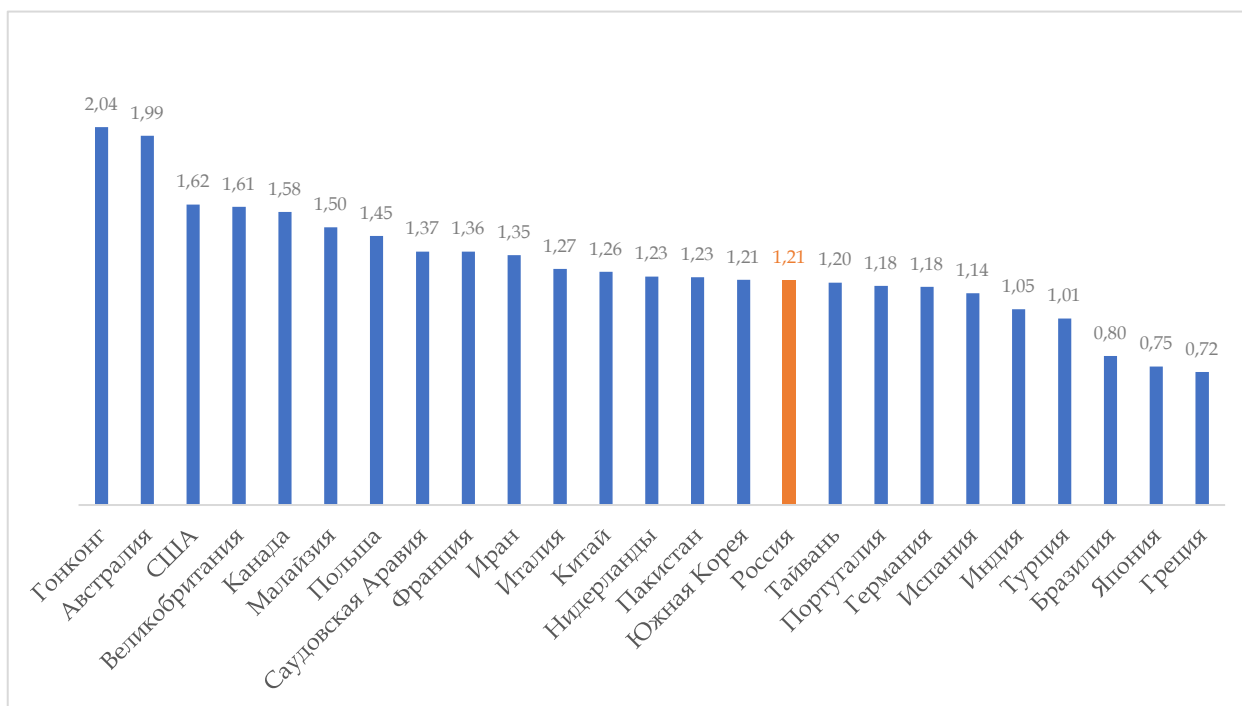


Рисунок 6. Средняя цитируемость публикаций, нормализованная по области исследований, году и типу публикаций (25 стран-лидеров по числу публикаций, 2019 г.)

Среди тенденций последних лет для пятерки самых крупных (по числу публикаций в области технологий работы с большими данными) стран и России можно отметить падение значений показателя у США, Великобритании и (наиболее выраженное) Германии; небольшой, но стабильный рост средней нормализованной цитируемости публикаций Китая, Индии и России (рис. 7).

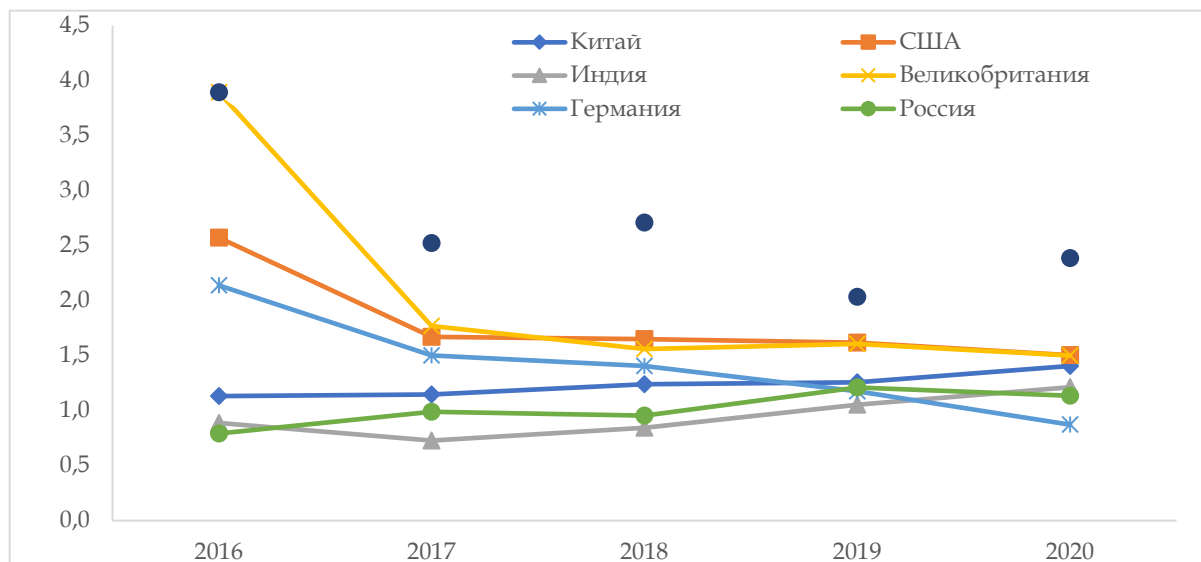


Рисунок 7. Динамика средней цитируемости публикаций, нормализованной по области исследований, году и типу публикаций (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.)

5.3 Международные и инновационные связи

Сотрудничество академических исследователей с корпоративным сектором является важным каналом трансфера технологий и индикатором инновационных связей на первых этапах инновационного цикла. Метрика этого сотрудничества («Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией») отражает потенциальную практическую значимость проводимых академических исследований и характеризует эффективность национальной инновационной системы – не случайно лидерами по этому показателю являются страны с высокой инновационной активностью предприятий и высоким уровнем технологического развития (рис. 8; из не вошедших в анализируемые 25 стран-лидеров по числу публикаций в области технологий работы с большими данными в этой связи можно отметить еще Израиль со значением этого показателя 10% и Швейцарию – со значением 14,81%). Россия по этому показателю в 2019 году занимала 19-е место среди 25 стран с наибольшим числом публикаций в области технологий работы с большими данными.

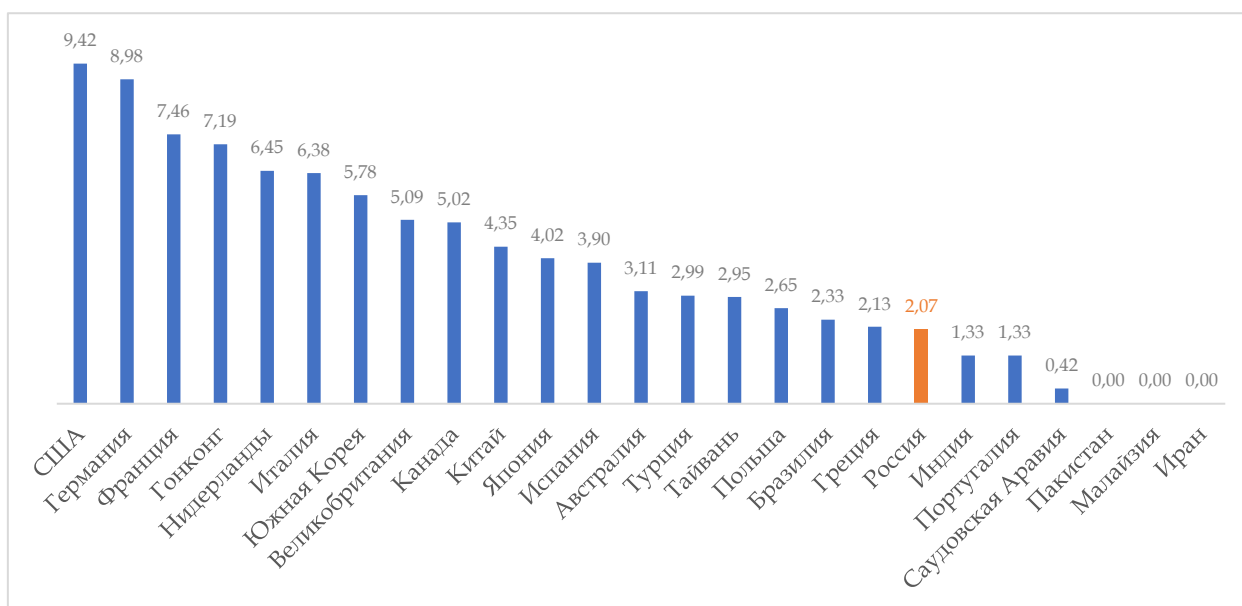


Рисунок 8. Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией (25 стран-лидеров по числу публикаций, 2019 г.), %

В последние годы среди пяти стран, наиболее активно публикующихся в области технологий работы с большими данными, Германия и США становились лидерами и по этому показателю (рис. 9). Среди анализируемых стран за последние 5 лет можно отметить небольшой тренд на снижение значения показателя у Германии, США, Великобритании и Китая и на увеличение сотрудничества с индустрией у Индии и России. При этом у России и Индии этот показатель остается ниже, чем у лидеров, а сами тенденции не являются выраженными.

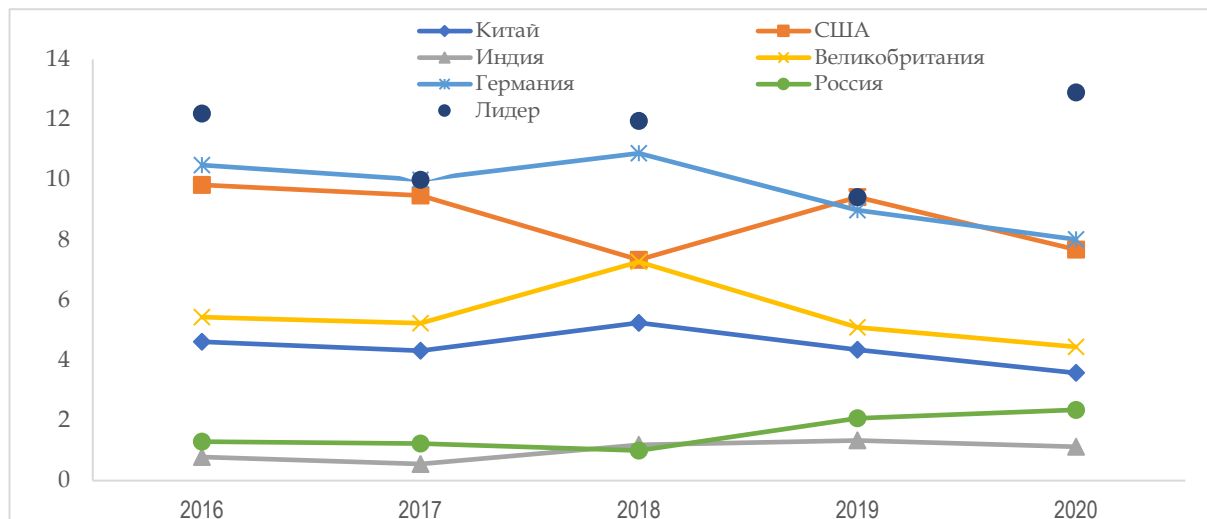


Рисунок 9. Динамика доли публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.), %

Международное научное сотрудничество — важный фактор развития национальных исследований и разработок. Совместная работа ученых стимулирует обмен научными знаниями и навыками исследования, позволяет получить доступ к уникальному научному оборудованию, весь спектр которого не может позволить себе ни одна из стран.

На рисунке 10 приведены данные о международном соавторстве публикаций 25 стран, представленных наибольшим числом публикаций в WoS. Россия занимает среди них 22-е место, публикуя в международном соавторстве около 31% работ.

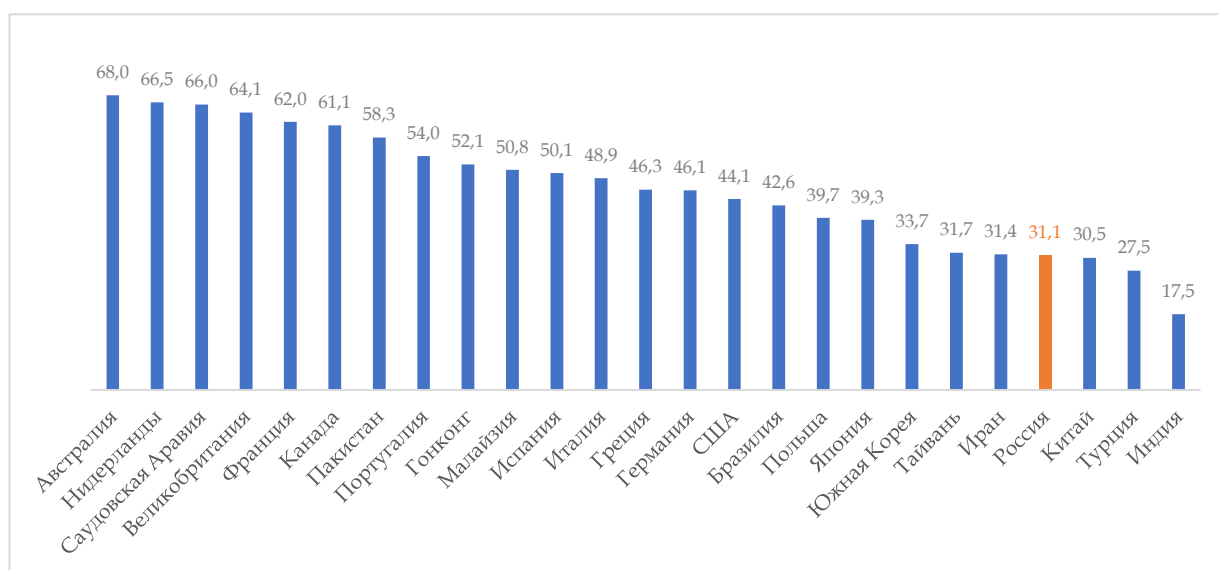


Рисунок 10. Доля публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве (25 стран-лидеров по числу публикаций, 2019 г.), %

Следует отметить, что интернационализация научной деятельности — одна из ярких и характерных тенденций развития современной науки. Рост масштабов и разнообразия форм международного научного сотрудничества существенно изменил за последние десятилетия облик науки. Эта тенденция — динамичный и устойчивый рост числа совместных публикаций исследователей из разных стран — отмечается в последние годы и для исследований в области технологий работы с большими данными (рис. 11). Россия в последние два года догнала Китай по числу совместных работ и опережает Индию. При интерпретации данного показателя следует принять во внимание, что исследователи из «больших» научных стран, как правило, меньше публикуются в международном соавторстве, находя партнеров внутри страны. Россия уступает по показателю международного сотрудничества почти всем странам, имеющим в WoS большее число публикаций, что является негативным симптомом.

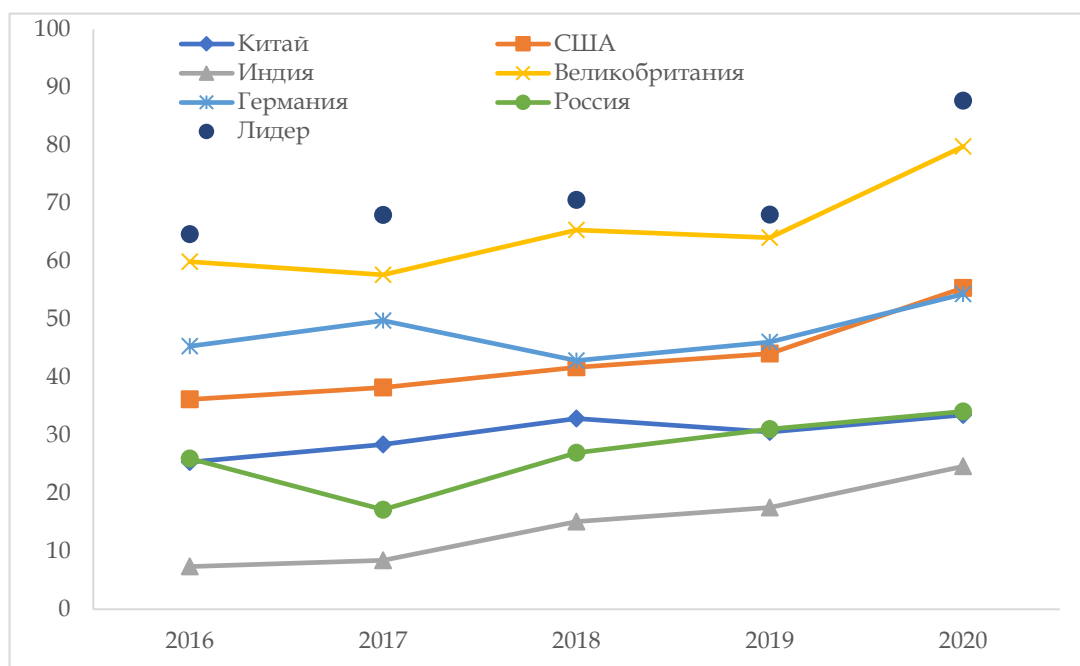


Рисунок 11. Динамика доли публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.), %

Заключение

По результатам пилотной реализации системы мониторинга исследований и разработок в области технологий работы с большими данными можно сделать несколько выводов:

- 1) разработанная методология адекватно отражает данную предметную область и может быть использована при проведении полномасштабных мониторинговых исследований;
- 2) предложенная концептуальная схема мониторинга и система показателей масштабируемы и позволяют включать в систему мониторинга (без изменения ее концептуальных рамок и методических основ) новые обследования и показатели;
- 3) опыт разработки и реализации системы мониторинга исследований и разработок в области технологий работы с большими данными может быть распространен на другие сквозные технологии цифровой экономики.

Полученные результаты свидетельствуют о том, что для российских исследований и разработок в области технологий работы с большими данными характерен ряд проблем:

- 1) масштабы публикационной активности российских исследователей, представленные в ведущих международных библиографических системах, не соответствуют масштабам экономики и научной деятельности в стране;

2) показатели уровня российских исследований близки к среднемировым, постепенно растут, но значительно уступают странам с наиболее масштабными исследованиями в области технологий работы с большими данными;

3) инновационные связи проводимых в стране исследований и разработок, их востребованность в коммерческом секторе остаются на низком уровне – Россия в разы (в отдельных случаях на порядок) уступает странам с развитой инновационной системой по показателям сотрудничества академических исследователей с коллегами из индустрии.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. ГОСТ Р 57194.1 – 2916 Трансфер технологий. Общие положения.
2. ГОСТ Р 58048 – 2017 Трансфер технологий. Методические указания по оценке уровня зрелости технологий.
3. OECD (2007) Revised Field of Science and Technology (FOS) Classification in the Frascati Manual. URL: <http://www.oecd.org/science/innovationinsciencetechnologyandindustry/38235147.pdf> (дата обращения 15.06.2021)
4. OECD Guide to Measuring the Information Society 2011. <https://doi.org/10.1787/9789264113541-en>
5. OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. <https://doi.org/10.1787/9789264239012-en>
6. Web of Science. Clarivate Analytics. URL: <https://csi.webofknowledge.com/CSI/exit.do?Func=Exit&SID=C1LIQAIG3E5mZMtoMhE> (дата обращения 15.06.2021)
7. Cornell University, INSEAD, and WIPO (2020). The Global Innovation Index 2020: Who Will Finance Innovation? URL: <https://www.globalinnovationindex.org/> (дата обращения 15.06.2021)
8. Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, “The AI Index 2021 Annual Report” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021. URL: <https://aiindex.stanford.edu/report/> (дата обращения 15.06.2021)
9. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (дата обращения 15.06.2021)
10. López Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A.-J., & Parra-González, M. E. (2020). Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science. *Symmetry*, 12(4), 495. <https://doi.org/10.3390/sym12040495>
11. Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. *Journal of Information Science*, 016555151878988. <https://doi.org/10.1177/0165551518789880>
12. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends. *J Big Data* 4, 30 (2017). <https://doi.org/10.1186/s40537-017-0088-1>

13. Liu, X., Sun, R., Wang, S. and Wu, Y.J. (2019). The research landscape of big data: a bibliometric analysis, *Library Hi Tech*, Vol. 38 No. 2, pp. 367-384. <https://doi.org/10.1108/LHT-01-2019-0024>
14. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis. *Scientometrics* 122, 1563–1581 (2020). <https://doi.org/10.1007/s11192-020-03371-2>
15. Bidosola, I., Garechana, G., Zarrabeitia, E. et al. Characterization of strategic emerging technologies: the case of big data. *Cent Eur J Oper Res* 28, 45–60 (2020). <https://doi.org/10.1007/s10100-018-0597-9>
16. Peng, Y., Shi, J., Fantinato, M. et al. A study on the author collaboration network in big data. *Inf Syst Front* 19, 1329–1342 (2017). <https://doi.org/10.1007/s10796-017-9771-1>
17. Parlina A, Ramli K, Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. *Information*. 2020; 11(2):69. <https://doi.org/10.3390/info11020069>
18. Ting-Peng Liang, Yu-Hsi Liu. Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, Volume 111, 2018, Pages 2-10. <https://doi.org/10.1016/j.eswa.2018.05.018>
19. Furlan, Patricia Kuzmenko; Laurindo, Fernando Jose Barbin (2017). Epistemological grouping of published articles on big data analytics. *TransInformacao*, 29(1), P. 91-100. <https://doi.org/10.1590/2318-08892017000100009>
20. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. *J Big Data* 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
21. Ahmad, I., Ahmed, G., Shah, S.A.A. et al. A decade of big data literature: analysis of trends in light of bibliometrics. *J Supercomput* 76, 3555–3571 (2020). <https://doi.org/10.1007/s11227-018-2714-x>
22. Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal Patent Analysis for Big Data Technology. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). doi:10.23919/picmet.2017.81254
23. Шапошник С.Б. Международное научное сотрудничество и публикационная активность российских ученых в Computer Science в 1993-2017 гг.: междисциплинарный и межстрановой сравнительный анализ // Информационное общество, 2018, №6, С. 39-45.
24. World Economic Outlook Database: April 2021. URL: <https://www.imf.org/en/Publications/WEO/weo-database/2021/April/weo-report?c=512,914,612,614,311,213,911,314,193,122,912,313,419,513,316,913,124,339,638,514,218,963,616,223,516,918,748,618,624,522,622,156,626,628,228,924,233,632,636,634,238,662,960,423,935,128,611,321,243,248,469,253,642,643,939,734,644,819,172,132,646,648,915,134,652,174,328,258,656,654,336,263,268,532,944,176,534,536,429,433,178,436,136,343,158,439,916,664,826,542,967,443,917,544,941,446,666,668,672,946,137,546,674,676,548,556,678,181,867,682,684,273,868,921,948,943,686,688,518,728,836,558,138,196,278,692,694,962,142,449,564,565,283,853,288,293,566,964,182,359,453,968,922,714,862,135,716,456,722,942,718,724,576,936,961,813,726,199,733,184,524,361,362,364,732,366,144,146,463,528,923,738,578,537,742,866,369,744,186,925,869,746,926,466,112,111,298,927,846,299,582,487,474,754,698,&s=PPPPC,&sy=2016&ey=2020&ssm=0&scsm=1&scs=0&ssd=1&ssc=0&sic=0&sort=countr y&ds=.&br=1> (дата обращения 15.06.2021)

RESEARCH AND DEVELOPMENT IN BIG DATA

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-Based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Shaposhnik, Sergei Borisovich

*Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher
Petrozavodsk, Russian Federation
sergei.shaposhnik@gmail.com*

Abstract

A conceptual framework and a set of indicators have been developed for monitoring R&D as a factor in the development and use of big data technologies. The framework has three main components: national scale of research and development, level of research and development in comparison with other countries, and international cooperation as well as cooperation between academia and business. The calculation of indicators and a comparative analysis of the results obtained with other countries were carried out.

Keywords

big data; big data technologies; research and development; publication activity; patent activity; scientometric indicators; monitoring and evaluation

References

1. GOST R 57194.1 – 2916 Transfer tekhnologiy. Obshchiye polozheniya.
2. GOST R 58048 – 2017 Transfer tekhnologiy. Metodicheskiye ukazaniya po otsenke urovnya zrelosti tekhnologiy.
3. OECD (2007) Revised Field of Science and Technology (FOS) Classification in the Frascati Manual. URL: <http://www.oecd.org/science/innovationinsciencetechnologyandindustry/38235147.pdf> (accessed on 15.06.2021)
4. OECD Guide to Measuring the Information Society 2011. <https://doi.org/10.1787/9789264113541-en>
5. OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. <https://doi.org/10.1787/9789264239012-en>
6. Web of Science. Clarivate Analytics. URL: <https://csi.webofknowledge.com/CSI/exit.do?Func=Exit&SID=C1LIQAIG3E5mZMtoMhE> (data obrashcheniya 15.06.2021)
7. Cornell University, INSEAD, and WIPO (2020). The Global Innovation Index 2020: Who Will Finance Innovation? URL: <https://www.globalinnovationindex.org/> (accessed on 15.06.2021)
8. Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, “The AI Index 2021 Annual Report” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021. URL: <https://aiindex.stanford.edu/report/> (accessed on 15.06.2021)
9. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (accessed on 15.06.2021)
10. López Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A.-J., & Parra-González, M. E. (2020). Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science. *Symmetry*, 12(4), 495. <https://doi.org/10.3390/sym12040495>
11. Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. *Journal of Information Science*, 016555151878988. <https://doi.org/10.1177/0165551518789880>

12. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends. *J Big Data* 4, 30 (2017). <https://doi.org/10.1186/s40537-017-0088-1>
13. Liu, X., Sun, R., Wang, S. and Wu, Y.J. (2019). The research landscape of big data: a bibliometric analysis, *Library Hi Tech*, Vol. 38 No. 2, pp. 367-384. <https://doi.org/10.1108/LHT-01-2019-0024>
14. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis. *Scientometrics* 122, 1563–1581 (2020). <https://doi.org/10.1007/s11192-020-03371-2>
15. Bidosola, I., Garechana, G., Zarrabeitia, E. et al. Characterization of strategic emerging technologies: the case of big data. *Cent Eur J Oper Res* 28, 45–60 (2020). <https://doi.org/10.1007/s10100-018-0597-9>
16. Peng, Y., Shi, J., Fantinato, M. et al. A study on the author collaboration network in big data. *Inf Syst Front* 19, 1329–1342 (2017). <https://doi.org/10.1007/s10796-017-9771-1>
17. Parlina A, Ramli K, Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. *Information*. 2020; 11(2):69. <https://doi.org/10.3390/info11020069>
18. Ting-Peng Liang, Yu-Hsi Liu Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, Volume 111, 2018, Pages 2-10. <https://doi.org/10.1016/j.eswa.2018.05.018>
19. Furlan, Patricia Kuzmenko; Laurindo, Fernando Jose Barbin (2017). Epistemological grouping of published articles on big data analytics. *TransInformacao*, 29(1), P. 91-100. <https://doi.org/10.1590/2318-08892017000100009>
20. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. *J Big Data* 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
21. Ahmad, I., Ahmed, G., Shah, S.A.A. et al. A decade of big data literature: analysis of trends in light of bibliometrics. *J Supercomput* 76, 3555–3571 (2020). <https://doi.org/10.1007/s11227-018-2714-x>
22. Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal Patent Analysis for Big Data Technology. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). doi:10.23919/picmet.2017.81254
23. Shaposhnik S.B. Mezhdunarodnoye nauchnoye sotrudnichestvo i publikatsionnaya aktivnost' rossiyskikh uchenykh v Computer Science v 1993-2017 gg.: mezhdistsiplinarnyy i mezhstranovoy sravnitel'nyy analiz // *Informatsionnoye obshchestvo*, 2018, №6, S. 39-45.
24. World Economic Outlook Database: April 2021. URL: <https://www.imf.org/en/Publications/WEO/weo-database/2021/April/weo-report?c=512,914,612,614,311,213,911,314,193,122,912,313,419,513,316,913,124,339,638,514,218,963,616,223,516,918,748,618,624,522,622,156,626,628,228,924,233,632,636,634,238,662,960,423,935,128,611,321,243,248,469,253,642,643,939,734,644,819,172,132,646,648,915,134,652,174,328,258,656,654,336,263,268,532,944,176,534,536,429,433,178,436,136,343,158,439,916,664,826,542,967,443,917,544,941,446,666,668,672,946,137,546,674,676,548,556,678,181,867,682,684,273,868,921,948,943,686,688,518,728,836,558,138,196,278,692,694,962,142,449,564,565,283,853,288,293,566,964,182,359,453,968,922,714,862,135,716,456,722,942,718,724,576,936,961,813,726,199,733,184,524,361,362,364,732,366,144,146,463,528,923,738,578,537,742,866,369,744,186,925,869,746,926,466,112,111,298,927,846,299,582,487,474,754,698,&s=PPPPC,&sy=2016&ey=2020&ssm=0&scsm=1&sc=0&ssd=1&ssc=0&sic=0&sort=country&ds=.&br=1> (accessed on 15.06.2021)

Информационное общество: политика и факторы развития

ЦИФРОВАЯ ИНФРАСТРУКТУРА ДЛЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Ершов Петр Сергеевич

*Институт развития информационного общества, руководитель Дирекции региональных программ
Москва, Российская Федерация
peter.erшов@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель Совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Разработаны концептуальная схема и набор показателей мониторинга и оценки цифровой инфраструктуры для работы с большими данными. Концептуальная схема состоит из трех компонентов: телекоммуникационная инфраструктура; вычислительная инфраструктура; платформенная инфраструктура. Приведены расчеты основных показателей для России и отдельных сфер деятельности, проведены международные сопоставления по состоянию на конец 2020 года. Отмечено существенное отставание России от стран ЕС по отдельным инфраструктурным показателям.

Ключевые слова

цифровая технология; большие данные; цифровая инфраструктура; информационная инфраструктура; телекоммуникационная инфраструктура; сеть передачи данных; вычислительная инфраструктура; цифровая платформа; мониторинг и оценка; Big Data for Digital Economy, BD4DE

Введение

Наличие развитой цифровой инфраструктуры, удовлетворяющей растущим потребностям в передаче, хранении и обработке больших данных, является одним из ключевых факторов, влияющих на развитие цифровой экономики в современном мире. Основу инфраструктуры для работы с большими данными составляют инфраструктура передачи данных на основе телекоммуникационных фиксированных и мобильных сетей, вычислительная инфраструктура на основе центров обработки данных, а также экосистемы цифровых платформ, предоставляющих богатый спектр цифровых сервисов для участников многосторонних рынков.

Сегодняшние потребности в высокоскоростной инфраструктуре передачи данных возросли настолько быстро, что глобальный интернет испытывает серьезные трудности, связанные не только с пандемией. По данным компании TeleGeography [1] в 2020 году средний международный интернет-трафик увеличился на 48%, в то время как пропускная способность выросла на 35%, что является самым большим ростом с 2013 года. В период с 2016 по 2020 годы средний и пиковый международный интернет-трафик увеличивался в среднем на 30% в год, тогда как среднегодовой темп роста пропускной способности составлял 29%. Эта тенденция во многом обусловлена необходимостью передачи больших массивов данных, для чего и нужна соответствующая инфраструктура как на глобальном уровне, так и внутри страны.

Быстро накапливающиеся объемы данных требуют наличия все большего количества вычислительных ресурсов. По оценкам компании Global Industry Analysts [2] объем мирового

© Ершов П.С., Хохлов Ю.Е., 2021

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_110

рынка услуг центров обработки данных будет расти начиная с 2020 года на 13,4%. Высокие темпы роста прогнозируются в Канаде (+11,8% в год), Германии (+10,6%) и Японии (+9,5%). При этом китайский рынок будет увеличиваться в объеме еще быстрее – на 17,5% ежегодно. Глобальными лидерами рынка услуг центров обработки данных по состоянию на май 2021 года являются транснациональные компании Amazon, Microsoft и Google, владеющие более 50% крупнейших центров обработки данных в мире. Страны-лидеры цифровой экономики активно инвестируют в развитие передовых ИКТ, особое место среди которых занимают высокопроизводительные вычисления (high performance computing), применяемые для параллельной обработки данных при решении сложных вычислительных задач. Прогнозируется, что данный сегмент рынка вычислительных ресурсов тоже будет иметь среднегодовой темп роста 5,5% [3].

Одним из ключевых драйверов цифровой трансформации экономики стали онлайн-цифровые платформы, появление которых принципиально изменило способы ведения бизнеса в отраслях [4]. Под цифровыми платформами, как правило, понимают реализованные с помощью цифровых технологий «многосторонние торговые площадки с бизнес-моделями, которые предоставляют производителям и потребителям возможность создавать обоюдовыгодные ценности, взаимодействуя друг с другом» [5]. Стремительный рост многосторонних цифровых платформ ведет к повышению производительности, снижению транзакционных издержек при взаимодействии, обеспечивает мгновенную и непрерывную связь между людьми, облегчает сотрудничество и кооперацию между организациями, стирая территориальные границы и преодолевая часовые пояса, а также усиливает конкурентоспособность национальных рынков [6], [7]. Цифровые платформы предоставляют новые возможности участникам рынка за счет взаимного влияния двух эффектов: сетевого и инновационного. Сетевой эффект возникает благодаря использованию цифровых технологий при осуществлении транзакций и объединяет миллионы потребителей на единой площадке. Инновационный эффект возникает там, где операторы платформ предоставляют в общее пользование данные и сервисы, способствуя тем самым совместным технологическим инновациям различных игроков рынка. В последние годы еще одним трендом в развитии стало формирование экосистем цифровых платформ, порождающих новые виды экономических отношений и еще больше усиливающих сетевые и инновационные эффекты за счет взаимной диффузии доселе отдельно функционирующих сегментов рынка [8].

Поэтому наличие цифровых платформ в стране или отдельной сфере деятельности является не менее важным фактором, влияющим на развитие цифровой экономики, чем наличие зрелой телекоммуникационной или вычислительной инфраструктуры, о которых шла речь выше.

В данной статье представлена концептуальная схема компонентов мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации (Big Data for Digital Economy, BD4DE) [9], относящихся к цифровой инфраструктуре для работы с большими данными в масштабах страны и в отдельных сферах деятельности, а также результаты проведенной на основе разработанной методологии (и части предложенных показателей) оценки этого фактора по данным в 2020 года.

1 Определение предметной области мониторинга и оценки

Предметом мониторинга в рамках данного исследования является состояние цифровой инфраструктуры, которое может как способствовать удовлетворению социальных и экономических потребностей основных заинтересованных сторон (граждан, бизнеса, власти, научно-образовательного сообщества) благодаря использованию цифровых технологий для работы с данными, так и выступать препятствием на пути развития информационного общества и цифровой экономики.

1.1 Обзор подходов к мониторингу и оценке цифровой инфраструктуры для работы с большими данными

Наблюдение и оценка состояния ИКТ-инфраструктуры как необходимого условия для развития информационного общества имеет давнюю историю и ведется как на международном уровне, прежде всего, Международным союзом электросвязи (далее – МСЭ) [10], так и на национальном уровне соответствующими администрациями связи и/или национальными статистическими ведомствами. За эти годы сложился общепризнанный перечень показателей, характеризующих в первую очередь уровень развития как телекоммуникационной инфраструктуры передачи данных,

так и вычислительной инфраструктуры для хранения и обработки данных. Эти наблюдения ведутся такими международными организациями, как МСЭ [11], Организация экономического сотрудничества и развития (далее – ОЭСР) [12], Евростат [13]. В Российской Федерации долгосрочное наблюдение по аналогичным показателям ведется Росстатом [14].

Использование традиционных показателей ИКТ-инфраструктуры для оценки цифровой инфраструктуры передачи, хранения и обработки больших данных нашло широкое распространение и применяется как для международных сопоставлений, так и между различными сферами деятельности. В то же время следует отметить, что для инфраструктурного уровня, связанного с активно формирующимися в последние годы на глобальном, национальном или отраслевом уровнях экосистемами цифровых платформ [15], [7] еще не разработаны устоявшиеся инструменты мониторинга и системы показателей, что приводит к трудностям при оценке и проведении международных или межотраслевых сопоставлений. В данной работе будет предложен авторский подход к мониторингу цифровых платформ, развивающий направление, обозначенное в работе [16], как к необходимому условию успешного развития и использования технологий работы с большими данными при решении социальных или экономических задач.

С учетом вышесказанного предметная область мониторинга цифровой инфраструктуры работы с большими данными должна содержать три основных компонента, позволяющих оценить наличие и доступность развитой телекоммуникационной инфраструктуры, наличие и доступность вычислительных ресурсов для хранения и обработки больших данных, наличие и доступность цифровых платформ для развития и использования технологий работы с большими данными.

1.2 Обзор литературы по мониторингу и оценке цифровой инфраструктуры для работы с большими данными

Для разработки концептуальной схемы мониторинга цифровой инфраструктуры для работы с большими данными библиометрическим методом был проведен анализ релевантных научных публикаций, индексируемых в платформе Web of Science (WoS).

На первом этапе из библиографической «ядерной» коллекции WoS Core Collection был выделен основной массив публикаций, связанных с большими данными на основе на основе следующего поискового образа:

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Подробности формирования приведенного выше поискового образа описаны в статье [9]. Поиск велся за период с 2016 по 2020 год по таким полям, как «Название», «Аннотация», «Автор» и «Ключевые слова» (применено поле «Тема» в инструментарии InCites платформы WoS), что в итоге дало в качестве результата немногим более 100 000 публикаций.

На следующем этапе из основного массива статей осуществлялся отбор публикаций, посвященных телекоммуникационной инфраструктуре, вычислительной инфраструктуре и инфраструктуре цифровых платформ. Для этого были сформированы специализированные поисковые образы:

- для телекоммуникационной инфраструктуры: “digital infrastructure” OR “information infrastructure” OR “ICT-infrastructure” OR “broadband” OR “broadband fixed” OR “broadband mobile”;
- для вычислительной инфраструктуры: “high-performance computing” OR “supercomputing” OR “cloud computing” OR “cloud service*” OR IaaS OR SaaS OR PaaS OR DaaS;
- для цифровых платформ: “digital platform*” OR “transaction platform*” OR “online platform*” OR “integrated platform*”.

Ключевые слова для уточняющих поисковых образов подбирались с целью максимально отразить исследуемую предметную (под)область и одновременно минимизировать информационный «шум» в форме публикаций, слабо связанных с исследуемой (под)областью.

Общий и специализированные поисковые запросы были объединены логическим оператором «AND», в результате поиска по полю «Тема» было найдено 1201 публикация за последние 5 лет. Распределение источников по годам показывает достаточно стабильный исследовательский интерес к цифровой инфраструктуре для работы с большими данными: за 2016 год найдено 167 публикаций, за 2017 г. – 226, за 2018 г. – 266, за 2019 г. – 283, за 2020 г. – 235.

На заключительном этапе был проведен углубленный анализ массива найденных публикаций: на первой стадии изучались аннотации, а затем в случае релевантности публикации – полный текст.

Телекоммуникационной инфраструктуре и широкополосному доступу для работы с большими данными посвящено 13 релевантных статей, в частности публикации, связанные с использованием широкополосного доступа в интернет (далее – ШПД) при работе с большими данными с помощью технологий интернета вещей [17, 18] и искусственного интеллекта [19, 20].

Вычислительной инфраструктуре для работы с большими данными, прежде всего, суперкомпьютерингу посвящено 20 публикаций, в том числе использованию высокопроизводительных вычислений для фундаментальных исследований или прогнозной аналитики [21, 22, 23, 24]. Особое место занимают публикации, посвященные сервисам облачных вычислений – общим числом 727, что демонстрирует широкую распространенность облачных сервисов при работе с большими данными.

Наконец, цифровым платформам для работы с большими данными посвящены 23 научные публикации, затрагивающие вопросы использования социальных сетей [25], анализа политических предпочтений пользователей [25, 26], изучения клиентского поведения [27], сферы образования [28, 29, 30, 31], здравоохранения [32, 33, 34, 35], интернета вещей [36, 37], культуры [38].

Анализ показал, что в отобранных публикациях при мониторинге телекоммуникационной инфраструктуры затрагиваются такие аспекты, как значимость фиксированного и мобильного ШПД для граждан и организаций, перспективы внедрения мобильной связи пятого поколения, а также экономическая доступность. Для развитых стран одной из серьезных проблем является ограниченная пропускная способность существующих каналов связи как внутри страны, так и межстрановых соединений.

При мониторинге вычислительной инфраструктуры для работы с большими данными исследователи делают акцент на потребности в хранении и обработке постоянно возрастающих объемов данных, а также на необходимости иметь в стране или отдельной сфере деятельности универсальные или специализированные вычислительные комплексы, такие как суперкомпьютеры или кластеры для параллельных вычислений.

В отношении оценки платформенной инфраструктуры консенсус по системе показателей для мониторинга пока не сформировался, однако большинство исследователей отмечает важность наличия цифровых платформ для трансформации практически во всех сферах и особенно при работе с большими данными. Важными трендами последних лет стало как появление специализированных цифровых платформ для конкретной сферы деятельности (здравоохранение, сельское хозяйство, транспорт, финансы и т.д.), так и объединение цифровых платформ в экосистемы и даже конкуренция цифровых экосистем между собой.

3 Концептуальная схема и показатели мониторинга

3.1 Концептуальная схема предметной области

Результаты проведенного анализа приводят к концептуальной схеме мониторинга цифровой инфраструктуры для работы с большими данными, состоящей из трех компонентов: телекоммуникационной инфраструктуры для передачи больших массивов данных, вычислительной инфраструктуры для хранения и обработки больших данных, платформенной инфраструктуры для оперирования большими данными (см. рисунок 1). Важно отметить, что выбор соответствующих компонентов подтверждается одним из первых исследований по анализу текущего состояния развития цифровой экономики в России [15], где соответствующие компоненты также составляют основу для мониторинга цифровой инфраструктуры.

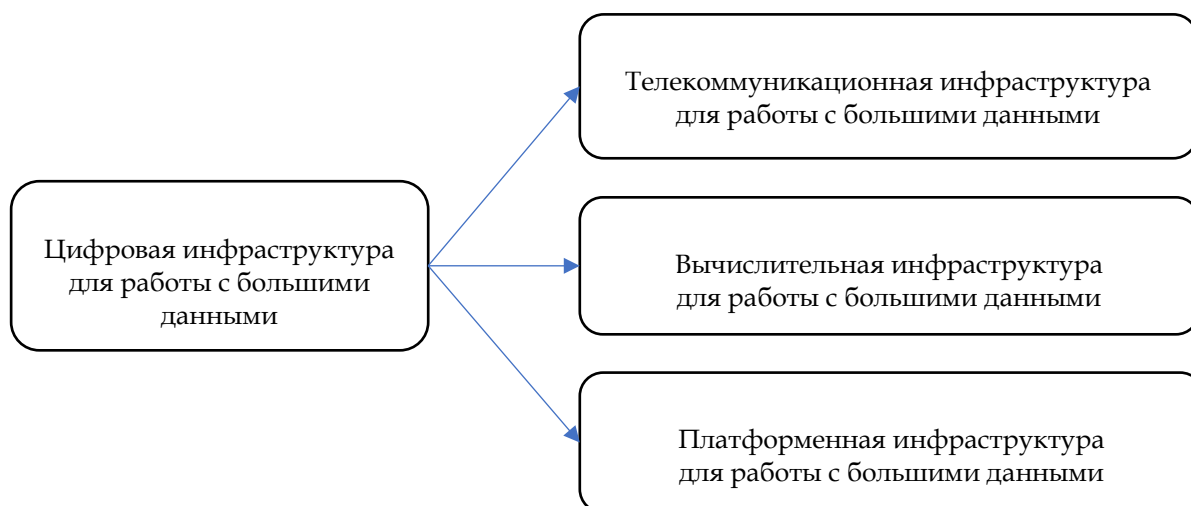


Рис. 1. Концептуальная схема мониторинга и оценки состояния цифровой инфраструктуры для работы с большими данными

Первый компонент предметной области мониторинга цифровой инфраструктуры для работы с большими данными связан с оценкой наличия телекоммуникационной инфраструктуры (страны, региона или сферы деятельности) и ее доступности для основных заинтересованных сторон. Важнейшими характеристиками телекоммуникационной инфраструктуры является скорость и полоса пропускания в сетях передачи данных, а также ее доступность, в том числе – экономическая. В рамках настоящего исследования критически важной является, прежде всего, такая характеристика телекоммуникационной инфраструктуры, как широкополосный доступ, обеспечивающий возможность использования инструментов распределенной работы с большими данными.

Не менее важным фактором, обеспечивающим возможность работы с большими данными, является наличие вычислительной инфраструктуры, предназначенной для хранения и обработки больших массивов данных и предоставляющей как вычислительные мощности центров обработки данных, так и разнообразный спектр сервисов облачных вычислений для конечных пользователей. Наличие в стране, регионе или конкретной сфере деятельности вычислительных ресурсов позволяет не только использовать современные технологии и инструменты работы с большими данными, но и поддерживать отечественных производителей и поставщиков цифровых решений и сервисов, формируя спрос в цифровом секторе экономики. При этом доступность (в том числе экономическая) вычислительных ресурсов для различных категорий пользователей является важной предпосылкой для использования технологий работы с большими данными.

Как уже отмечалось, последние десятилетия характеризуются появлением и доминированием онлайн-цифровых платформ, ставших одним из главных драйверов цифровой трансформации во многих сферах деятельности. При этом сформировалось несколько направлений развития цифровых платформ, связанных как со способами ведения бизнеса, так и со стеком используемых в платформе цифровых технологий [4]. Для целей настоящего исследования существенной является оценка наличия и доступности в стране или конкретной сфере деятельности одной или нескольких цифровых платформ, позволяющих принципиально изменить способ ведения бизнеса и использовать большие данные.

Мониторинг цифровой инфраструктуры для работы с большими данными представляет интерес не только для международных сопоставлений на уровне стран, но и для оценки ситуации в отдельных сферах деятельности. В рамках настоящего исследования рассматриваются приоритетные сферы деятельности, связанные с национальными целями развития Российской Федерации до 2030 года [39]:

- 1) государственное управление;
- 2) здравоохранение;
- 3) наука;

- 4) образование;
- 5) промышленность;
- 6) развитие городской среды;
- 7) сельское хозяйство;
- 8) строительство;
- 9) транспорт и логистика;
- 10) финансовые услуги;
- 11) экология и природопользование;
- 12) энергетическая инфраструктура.

3.2 Телекоммуникационная инфраструктура для работы с большими данными

Первый компонент системы мониторинга цифровой инфраструктуры для работы с большими данными должен обеспечить возможность оценки наличия телекоммуникационной инфраструктуры и ее доступности, прежде всего, ШПД. Для целей настоящего исследования будут в первую очередь использоваться согласованные на международном уровне показатели, характеризующие ШПД для организаций, а также экономическую доступность подобных соединений [40].

Не менее важно оценить наличие в стране надежной телекоммуникационной инфраструктуры для использования гражданами, которые, с одной стороны, выступают в качестве потенциальных клиентов для бизнеса, использующего технологии работы с большими данными, а с другой стороны – производят большое количество пользовательских данных, необходимых для разработки и предоставления персонализированных сервисов. Для эффективного решения задач подобного рода необходимы развитая телекоммуникационная инфраструктура сбора и передачи данных, равно как и повсеместный доступ граждан к таким сервисам с другой.

Первую группу характеристик образуют показатели наличия телекоммуникационной инфраструктуры в стране.

- (ТИБД-01-01) Доля организаций, имеющих ШПД к интернету, в общем числе обследованных организаций

Данный показатель характеризует уровень зрелости телекоммуникационной инфраструктуры широкополосного (как фиксированного, так и мобильного) доступа к интернету, что необходимо для массового использования технологий работы с большими данными.

- (ТИБД-01-02) Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, в общем числе обследованных организаций

Высокие скорости доступа к интернету дают преимущества организациям перед конкурентами с меньшей скоростью, так как они имеют возможность быстрее получать, обрабатывать и извлекать «добавленную ценность» из больших массивов данных.

- (ТИБД-01-03) Доля населения, охваченного услугами подвижной связи не ниже 3G

При расчете показателя учитывается население, охваченное услугами сотовой связи 3-го и 4-го поколений. Основные области применения технологий для работы с большими данными – пользовательские данные, которые в основном генерируются мобильными устройствами пользователей

- (ТИБД-01-04) Доля населения, использующего мобильные устройства для доступа к интернету

Распространенность мобильных устройств (таких как смартфоны и планшеты) служат дополнительной характеристикой к показателю ТИБД-01-03, позволяющей оценить возможности для генерации, сбора и обработки пользовательских данных, которых лишены обычные мобильные телефоны.

- (ТИБД-01-05) Международная пропускная способность каналов связи на 1 пользователя

Данный показатель важен в силу того, что совокупная пропускная способность всех международных каналов связи является ограниченной величиной и его усредненная величина на каждого пользователя показывает скорость связи пользователя с внешним миром.

- (ТИБД-01-06) Стоимость корзины услуг фиксированного ШПД к интернету

Доступность ШПД заключается не только в наличии каналов для передачи данных, но и в экономической возможности населения оплачивать доступ в интернет. Поэтому МСЭ уже

достаточно длительное время мониторит стоимость корзины услуг фиксированного ШПД к интернету как плату за ежемесячное подключение к интернету (с ограничением объема трафика в 5 Гб), рассчитанную как доля от среднемесячного валового национального дохода (ВНД) страны на душу населения.

- (ТИБД-01-07) *Стоимость корзины услуг мобильного ШПД к интернету*

Аналогичный предыдущему показатель экономической доступности мобильного ШПД отражает стоимость ежемесячного использования мобильной связи с активным использованием мобильного (безлимитного) интернета, рассчитанную как доля от среднемесячного ВНД страны на душу населения.

3.3 Вычислительная инфраструктура для работы с большими данными

Как отмечалось ранее, работа с большими данными требует наличия высокопроизводительных вычислительных систем, способных обеспечить хранение и обработку больших массивов разнообразных данных, генерируемых (или собираемых) с большой скоростью. До недавнего времени для этого хватало традиционных ИКТ-решений, однако быстрое накопление огромных массивов цифровых данных в последнее десятилетие привело к необходимости создания и применения высокопроизводительных систем с принципиально другой архитектурой и производительностью. Наличие в стране (или отдельной сфере деятельности) подобных вычислительных ресурсов и сервисов на их основе открывает возможности для высокопроизводительной аналитики больших данных и моделирования сложных процессов.

Владение подобными высокопроизводительными вычислительными системами предполагает существенные расходы на содержание физической инфраструктуры центров обработки данных с большим количеством компьютерного оборудования, расходы на электроэнергию, обеспечение информационной и физической безопасности, бесперебойности и многое другое. Это далеко не всегда экономически выгодно даже для очень крупных компаний, поэтому выходом может стать совместное использование вычислительных ресурсов, предоставляемых по модели облачных вычислений. Поэтому в систему мониторинга вычислительной инфраструктуры для работы с большими данными включены не только показатели наличия соответствующих вычислительных мощностей в стране, но и характеристики сервисов облачных вычислений.

- (ВИБД-02-01) *Суммарная доля мощностей суперкомпьютеров страны в мировом рейтинге TOP500*

Для оценки уровня вычислительных ресурсов выбрана общая вычислительная мощность всех суперкомпьютеров страны из мирового рейтинга TOP500, отнесенная к размеру экономики, что косвенно характеризует обеспеченность национальной экономики вычислительными ресурсами, в том числе для работы с большими данными.

- (ВИБД-02-02) *Мощность коммерческих центров обработки данных*

В бизнес-среде достаточно распространенным показателем развитости рынка услуг центров обработки данных служит количество стойко-мест в отдельно взятом центре. При всей неоднозначности данного показателя он, тем не менее, может служить качественной характеристикой наличия в стране или отдельно взятой сфере деятельности вычислительных ресурсов для работы с большими данными.

- (ВИБД-02-03) *Доля объема рынка облачных вычислений от мирового*

Данный показатель в большей степени характеризует возможности национальной системы облачных вычислений и экономическую доступность облачных сервисов.

- (ВИБД-02-04) *Доля организаций, использующих сервисы облачных вычислений в общем числе обследованных организаций*

Востребованность и доступность сервисов облачных вычислений, особенно в разрезе по видам облачных сервисов (IaaS, PaaS, SaaS, DaaS и т.д.), открывают возможность для мониторинга и оценки облачной инфраструктуры как в масштабах страны, так и отдельной сферы деятельности.

Данные для показателя ВИБД-02-01 публикуются дважды в год на веб-сайте рейтинга TOP 500 [41]; сведения для расчета показателей ВИБД-02-02 и ВИБД-02-03 регулярно публикуются авторитетными консалтинговыми компаниями (см. например [3], [42] и могут быть рассчитаны в ходе кабинетного исследования; данные для расчета показателя ВИБД-02-04 ежегодно публикуются

по итогам официального статистического наблюдения как в Российской Федерации [43], так и в Европейском Союзе [44], что позволяет проводить международные сопоставления.

3.4 Платформенная инфраструктура для работы с большими данными

Современные цифровые платформы построены на работе с (большими) данными как потребителей и поставщиков, так и технологических операторов. Объединение разнообразных цифровых платформ в экосистемы еще в большей степени способствует взаимодействию между различными группами пользователей, предоставляя цифровые сервисы для проведения транзакций, сбора, обработки и совместного использования данных, связанных с их общими интересами или деятельностью.

Для оценки платформенной инфраструктуры в стране будут использованы следующие группы показателей: наличие универсальных цифровых платформ (например, платформ электронной коммерции, поисковых систем, контент-платформ, социальных сетей); наличие специализированных цифровых платформ для отдельных сфер деятельности; наличие технологических платформ, позволяющих использовать тот или иной набор цифровых технологий для работы с данными (например, машинного обучения или интернета вещей).

- (ПИБД-03-01) *Наличие доступных универсальных цифровых платформ*

Данный показатель характеризует наличие доступа пользователей (в том числе экономически приемлемого) к сервисам универсальных цифровых платформ как глобальных (при условии их локализации), так и национальных. При этом успешное функционирование подобных платформ, с одной стороны, порождает огромные массивы пользовательских данных, а с другой – требует новых методов их обработки, выступая драйвером развития технологий работы с большими данными.

- (ПИБД-03-02) *Наличие доступных специализированных цифровых платформ для отдельных сфер деятельности*

Включение данного показателя в систему мониторинга и оценки платформенной инфраструктуры открывает возможность для сравнительного анализа ситуации в разрезе по различным сферам деятельности как внутри страны, так и для международных сопоставлений.

- (ПИБД-03-03) *Наличие технологических цифровых платформ*

Показатель нацелен на мониторинг и оценку уровня технологического развития цифровых платформ в стране и для проведения международных сопоставлений. Его также можно рассматривать в разрезе отдельных сфер деятельности и оценивать их технологическую оснащенность для работы с данными.

- (ПИБД-03-04) *Наличие экосистемы цифровых платформ*

Использование данного показателя предоставит возможности мониторинга дальнейшей эволюции платформенной инфраструктуры, прежде всего, с точки зрения проникновения единых экосистем в различные сферы деятельности, равно как и объединения в единой экосистеме стека разнообразных цифровых технологий работы с данными. При этом важно отслеживать не только наличие экосистем в стране, но и их доступность (включая экономическую) и уровень конкуренции.

Дальнейшее развитие предложенной системы показателей для платформенной инфраструктуры может идти в направлении учета аспектов, связанных с повышением уровня зрелости как универсальных/специализированных платформ, так и технологических цифровых платформ. Аналогичные показатели можно ввести и по отношению к экосистемам цифровых платформ.

4 Результаты пилотной реализации

В настоящее время многие показатели мониторинга цифровой инфраструктуры для работы с большими данными можно рассчитать методом кабинетных исследований на основе открытых данных официальных статистических наблюдений. Ниже приведены результаты, основанные на последних доступных данных по итогам 2019 года. В ряде случаев показатели рассчитаны для Российской Федерации в целом, а в некоторых случаях – для отдельно взятых сфер деятельности. Всюду, где возможно, даются международные сопоставления Российской Федерации со странами Европейского союза.

4.1 Телекоммуникационная инфраструктура для работы с большими данными

В 2019 году по показателю «ТИБД-01-01 Доля организаций, подключенных к интернету, в общем числе обследованных организаций» Россия достигла значения 86,6%, что соответствует уровню таких европейских стран, как Венгрия – 86,88% и Латвия 87,91%, меньше значение только у Греции: 81,55% (см. ниже рисунок 2).

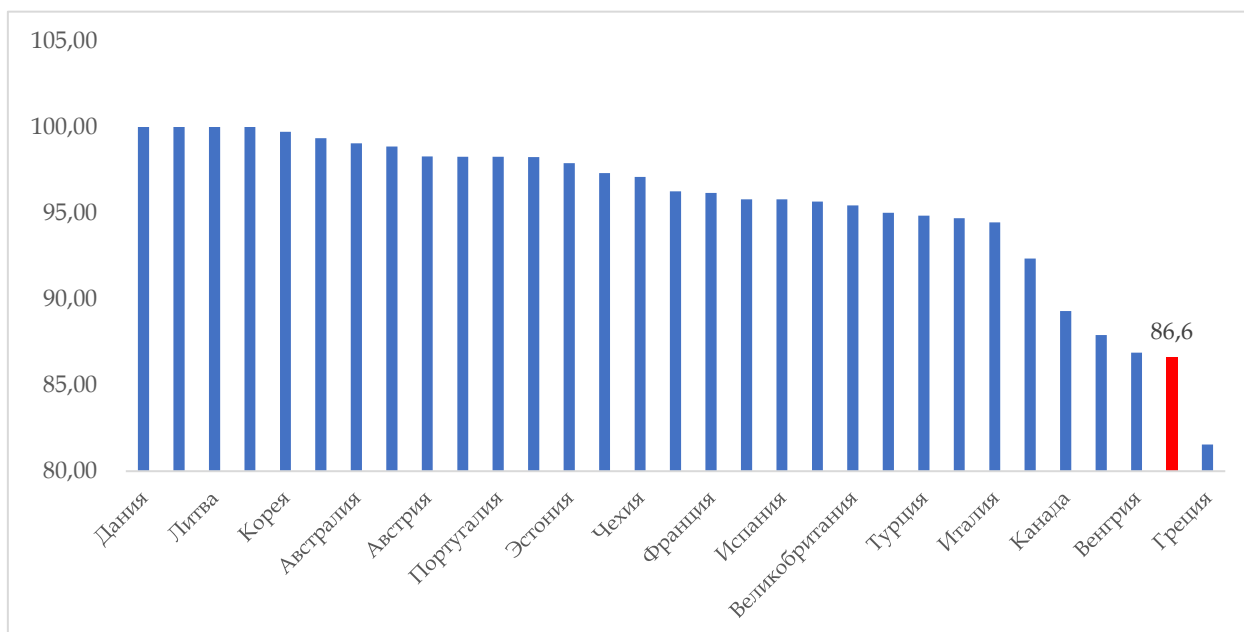


Рис. 2. Доля организаций, подключенных к интернету, %, 2019, ЕС, Россия

Источники: [43] [44]

Следует отметить, что большинство европейских стран перешагнуло 80%-й рубеж еще в 2010 году, а в 2019 среднее значение для стран ЕС составляет 97% и остается неизменным в течение последних пяти лет. В России идет замедляющийся рост количества организаций, имеющих доступ к интернету: например, с 2018 по 2019 год прирост составил всего 0,1%.

Наряду со страновыми показателями интернет-коннективности российских организаций, представляет интерес оценка по этому показателю для различных сфер деятельности (см. рисунок 3).

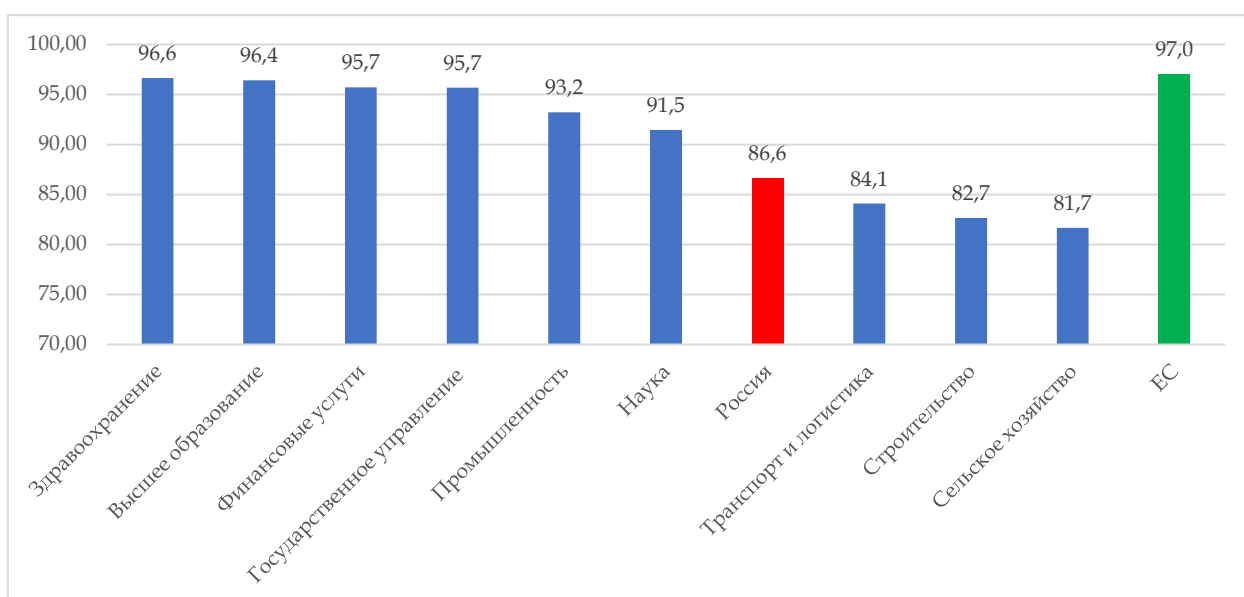


Рис. 3. Доля организаций, подключенных к интернету, %, 2019, в разрезе сфер деятельности,

Источники: [43] [44]

Отраслями-лидерами в Российской Федерации в 2019 году по показателю ТИБД-01-01 являются здравоохранение, высшее образование, финансы, государственное управление, промышленность и наука, для которых значение данного показателя выше среднего значения по России. Для транспорта, строительства и сельского хозяйства значения показателя – ниже среднего. При этом лидирующие сферы деятельности не сильно уступают среднему значению для Европейского союза. Для таких сфер, как развитие городской среды и энергетическая инфраструктура, данные в открытых источниках не доступны.

По показателю (ТИБД-01-02) «Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек» (см. рисунок 4) Россия существенно уступает странам ЕС обеспечивая в 2019 году ШПД на скорости не менее 30 Мбит/сек примерно 34% организаций, тогда как страны-лидеры достигли значений более 60% (Дания, Швеция, Португалия, Люксембург, Финляндия).

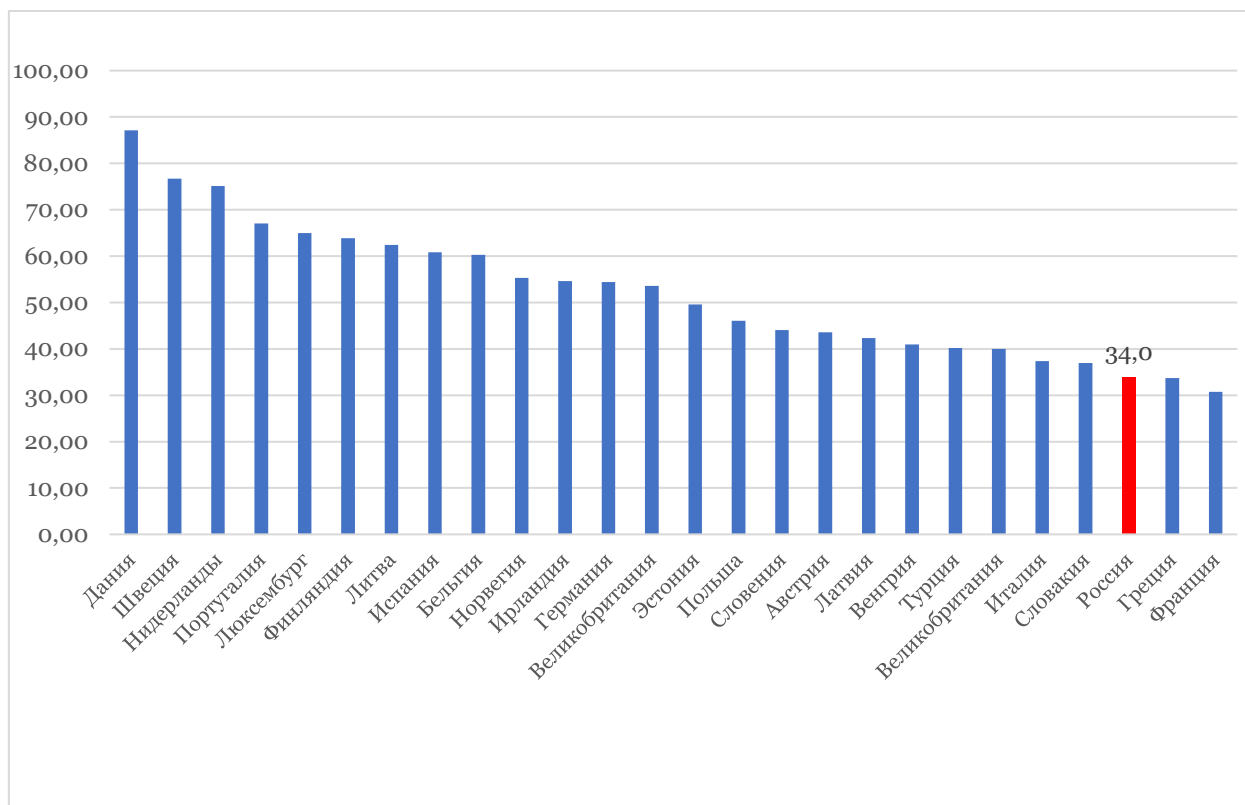


Рис. 4. Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, %, 2019, ЕС, Россия

Источники: [43] [44]

Динамика изменения показателя (ТИБД-01-02) последние 5 лет демонстрирует устойчивый рост для всех стран без исключения. Страны-лидеры ЕС при этом в среднем прибавляют в год по 5 процентных пунктов, что более чем в два раза превышает рост в Российской Федерации. Такими темпами европейские страны-лидеры решат задачу обеспечения организаций телекоммуникационной инфраструктурой в ближайшие 5–7 лет, тогда как России при сохранении имеющихся темпов для достижения сопоставимого уровня понадобится 30–40 лет.

В то же время ситуация с ШПД для организаций из отдельных сфер деятельности сильно различается: высшее образование, наука, промышленность и здравоохранение существенно превышают средние российские значения, тогда как система государственного управления и сельское хозяйство значительно уступают (см. рисунок 5). Особо следует отметить сферу высшего образования, где ШПД заметно превышает даже общеевропейский уровень.

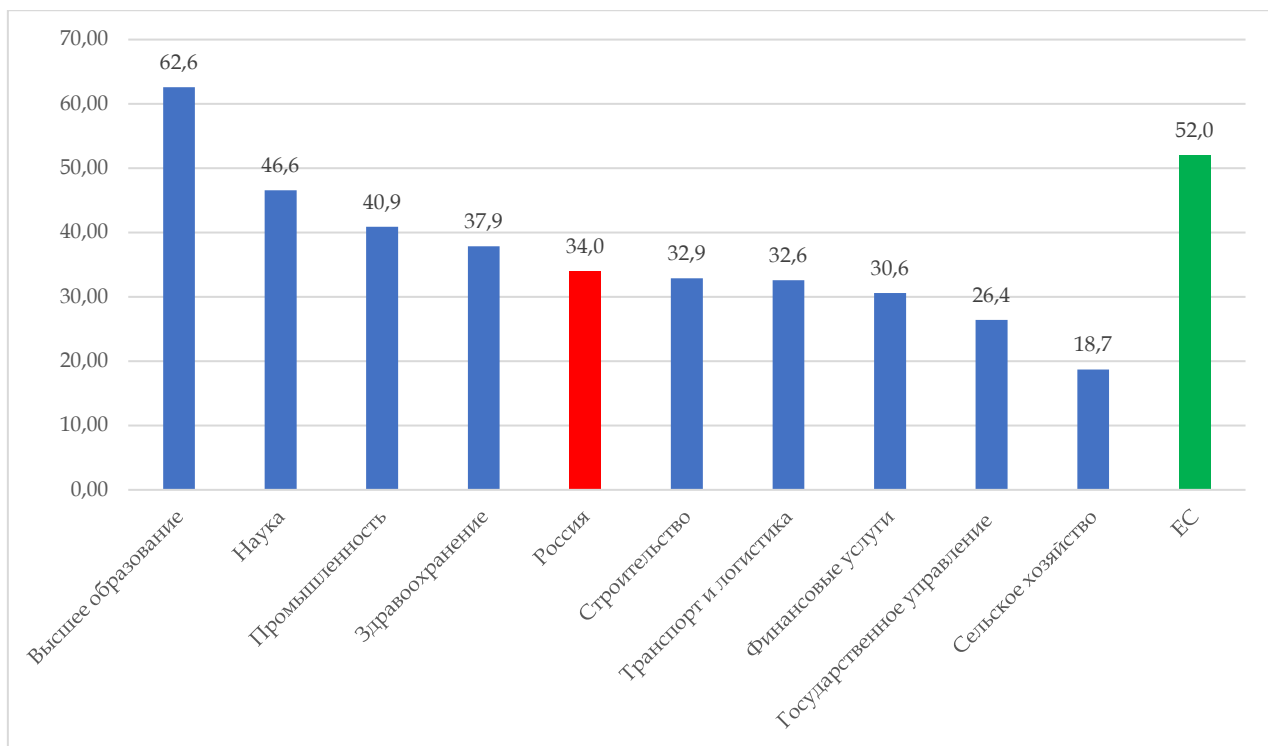


Рис. 5. Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, %, 2019, в разрезе сфер деятельности, Источники: [43] [44]

Охват населения услугами подвижной связи третьего поколения и выше характеризуется показателем ТИБД-01-03, по которому в 2019 году Россия достигла значения 87,7% (см. рисунок 6), тогда как для стран или страновых объединений, сопоставимых с Россией по размерам (ЕС, США, Китай, Бразилия, Канада), эта задача практически полностью решена (разве что кроме Бразилии).

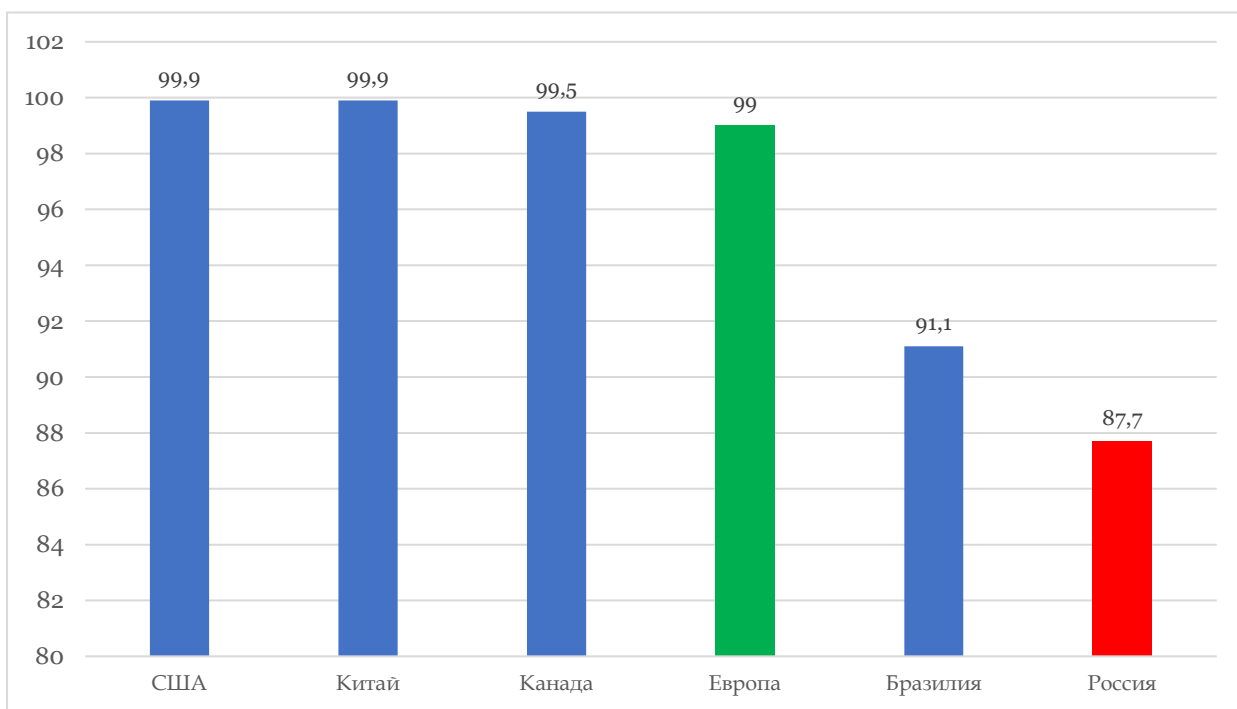


Рис. 6. Охват населения услугами подвижной связи не ниже 3G, %, 2019, Источники: [45]

В 2018 году 73,5% населения России использовало телефон или смартфон для доступа к интернету, что соответствует среднему уровню ЕС – 72% (см. рисунок 7), при этом к 2020 году наблюдалось снижение этих показателей [46].

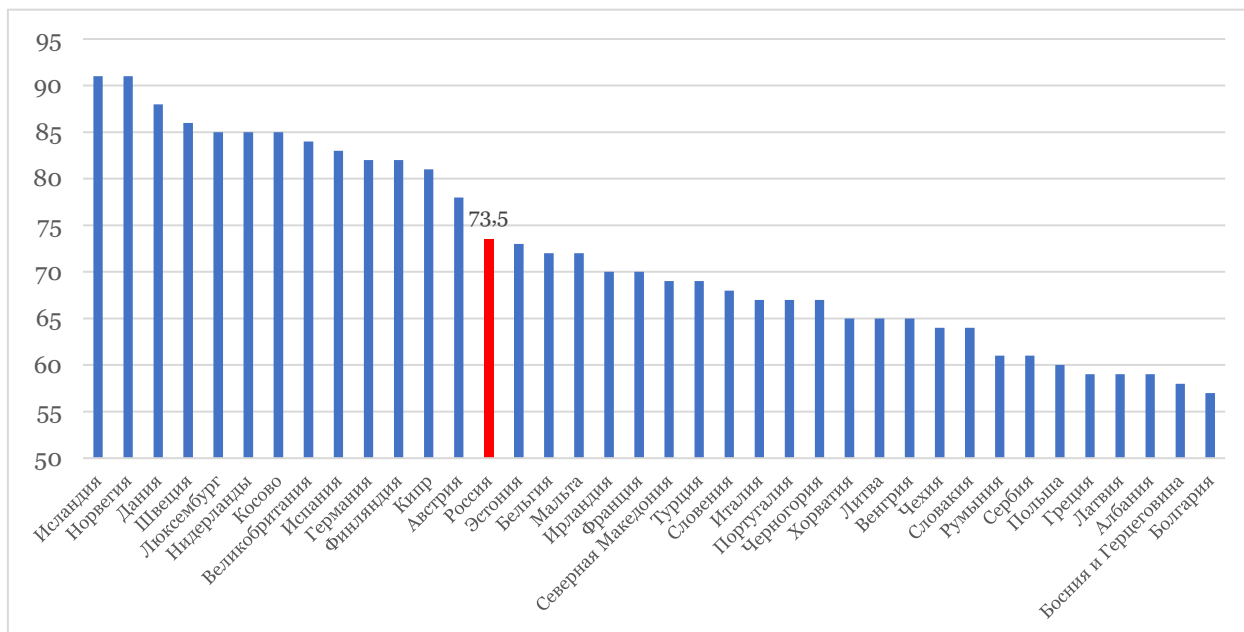


Рис. 7. Использование населением телефона или смартфона для доступа к интернету, %, 2018, Россия, ЕС
Источники: [44], [46]

По показателю (ТИБД-01-05) «Международная пропускная способность каналов связи на 1 пользователя» в 2019 году РФ со значением в 62 842,6 Кбит/с находится в конце рейтинга, опережая такие страны, как Тунис, Казахстан, Монголия, Египет и Куба. Лидерами являются Гонконг (11 614 411 Кбит/с) и Кипр (1 178 121 Кбит/с). Ввиду более чем 180-кратного превышения значения лидера (Гонконга) относительно России, на графике он не приводится (см. рисунок 8).

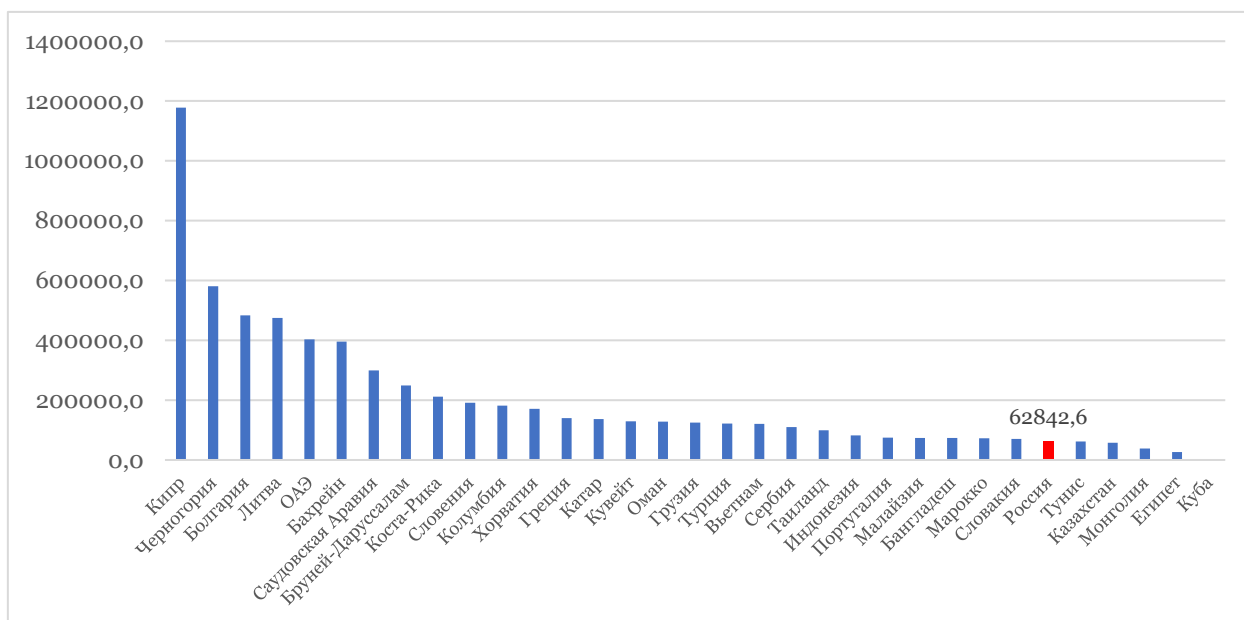


Рис. 8. Международная пропускная способность каналов связи на 1 пользователя
Источники: [45]

Динамика изменения международной пропускной способности стран показывает, что с 2015 по 2019 год страны-лидеры показывали значительный рост, кроме Болгарии в 2018 году. Гонконг в среднем за год прибавлял по 1 357 801,8 Кбит/с, Кипр – более чем 220 000 Кбит/с, другие европейские страны-лидеры – от 66 000 Кбит/с (Болгария и Литва) до 97 000 Кбит/с (Черногория).

Россия существенно отстает по темпам роста, с 2015 по 2017 год демонстрируя рост, который затем сменился спадом в 2018 и 2019 годах на 4 000 и 2 000 Кбит/с соответственно.

По показателям экономической доступности в 2020 году Россия вошла в десятку стран-лидеров по стоимости услуг фиксированного ШПД и заняла 8-е место со значением в 0,64% (ВНД на душу населения) из 178 стран (см. рисунок 9). Лидерами рейтинга являются Лихтенштейн, Кувейт, Китай и Объединенные Арабские Эмираты; аутсайдерами рейтинга являются Малави, Нигер и Мадагаскар, где стоимость услуг фиксированного ШПД превышает 100% ВНД на душу населения.

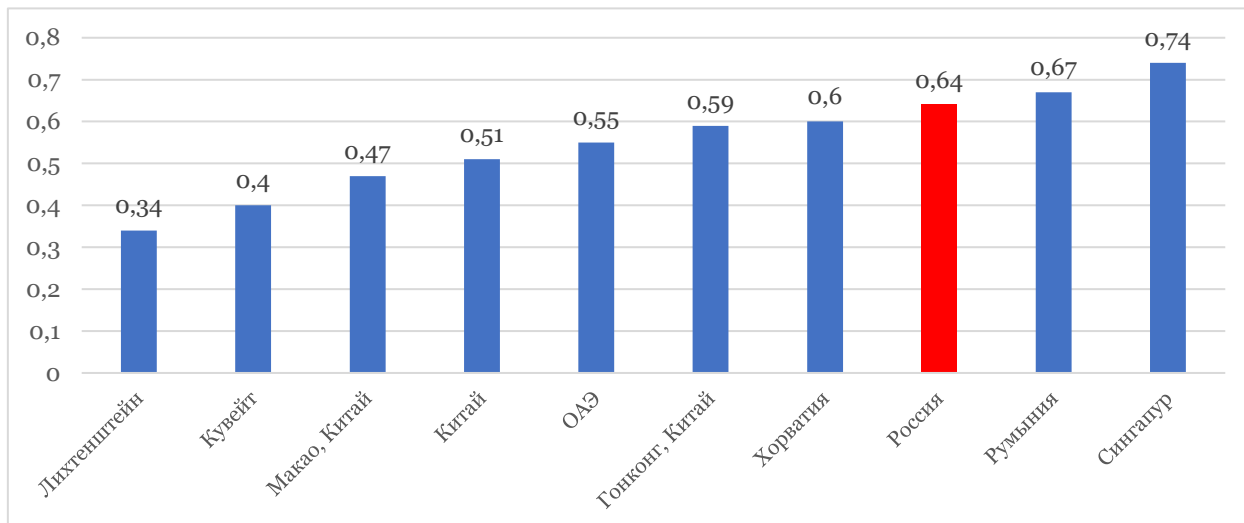


Рис. 9. Стоимость фиксированного ШПД в 2020 г., % ВНД на душу населения
Источники: [47]

По стоимости услуг мобильного ШПД в 2020 году Россия входит в двадцатку стран-лидеров и занимает 19-е место со значением в 0,67% (ВНД на душу населения) из 178 стран (см. рисунок 10). Лидерами являются Лихтенштейн, Люксембург, Австрия, Китай и Израиль со значениями до 0,3% (ВНД на душу населения); хуже всего дело обстоит в таких африканских странах, как Чад, Центральноафриканская Республика, Бурунди, Малави и Демократическая Республика Конго, где стоимость услуг мобильного ШПД превышает 50% (ВНД на душу населения).

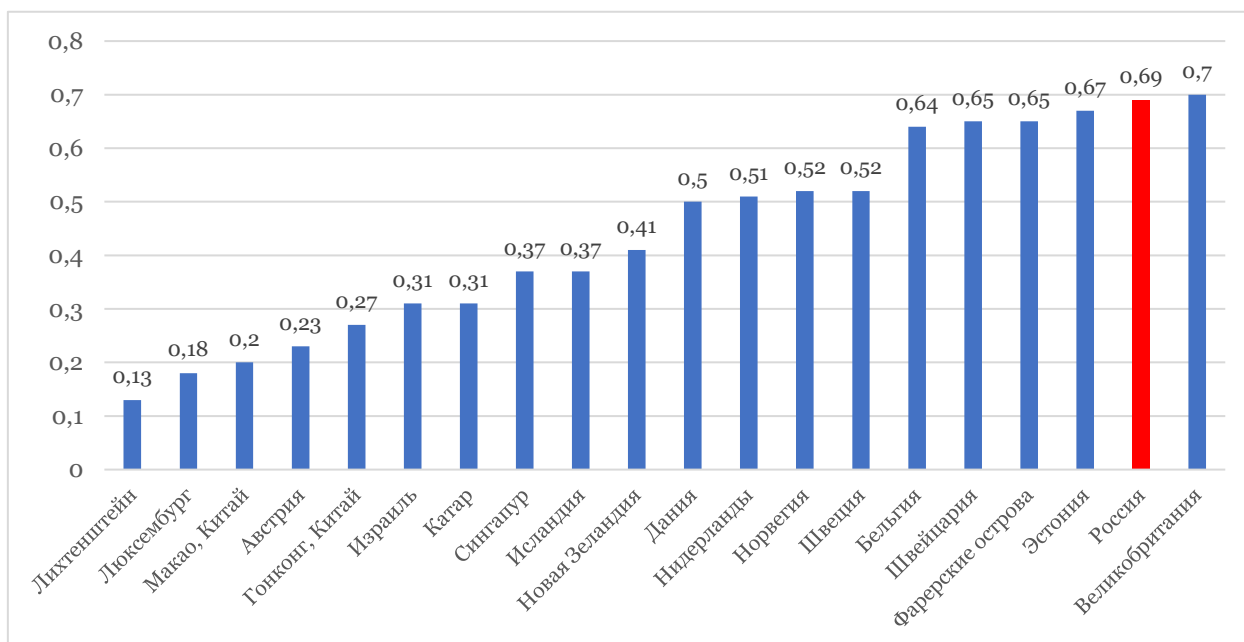


Рис. 10. Стоимость мобильного ШПД в 2020 г., % ВНД на душу населения
Источники: [47]

4.2 Вычислительная инфраструктура для работы с большими данными

Как уже отмечалось в разделе 3.3, одним из показателей наличия в стране вычислительных ресурсов для работы с большими данными служит суммарная доля мощностей суперкомпьютеров в рейтинге TOP 500 (показатель ВИБД-02-01).

Анализ рейтинга показал, что в 2019 году доля России по этому показателю составляла всего 0,63% что соответствует вычислительной мощности в 10347,3 Терафлопс. В 2020 году по сравнению с 2019 годом значение показателя ВИБД-02-01 еще уменьшилось и в рейтинге TOP 500 осталось всего 2 российских суперкомпьютера с вычислительной мощностью 9147 Терафлопс. Лидером рейтинга по числу компьютеров в ноябре 2020 являлся Китай (212 систем в списке), на втором месте США (113 систем), далее Япония (34 системы). При этом, несмотря на меньшее количество суперкомпьютерных систем, США возглавляют список по совокупной производительности в 668,7 Петафлопс (27,5% от суммарной производительности всех систем рейтинга), у Китая – 566,0 Петафлопс (23,2%). Благодаря лидирующей системе Fugaku Япония с общей производительностью суперкомпьютеров в 593,7 Петафлопс (24,4%) опережает Китай и находится на втором месте в мире (см. рисунок 11).

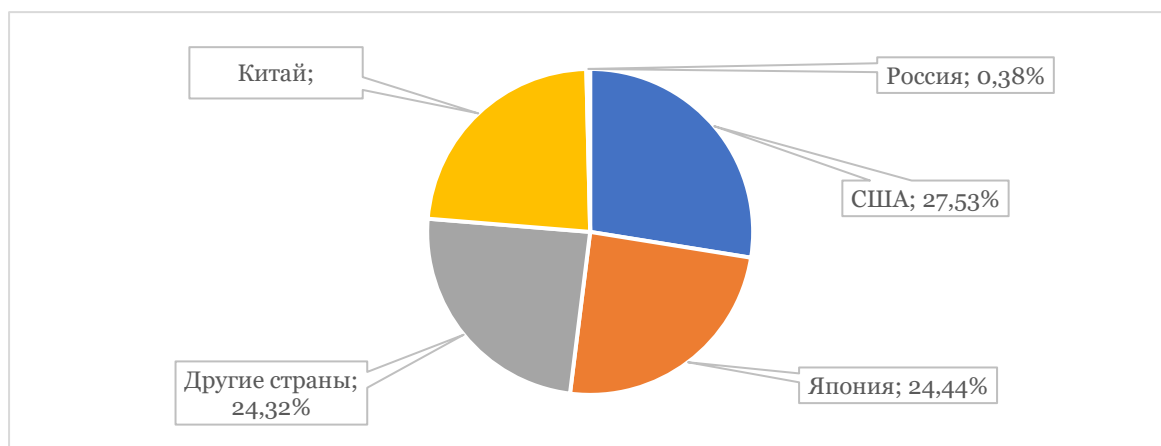


Рис. 11. Доля вычислительной мощности в рейтинге ТОП500 лидеров и России, %, 2020
Источники: [41]

По использованию сервисов облачных вычислений организациями (показатель ВИБД-02-04) в 2018 году Российская Федерация достигла значения в 26%, что соответствует среднему уровню ЕС (26%), лидерами являются скандинавские страны Финляндия – 65%, Швеция – 57%, Дания – 56%, Норвегия – 51% и Нидерланды – 48% (см. рисунок 12). В 2019 году значение Российской Федерации по данному показателю выросло до 28%, в ЕС данные за этот год не собирались.

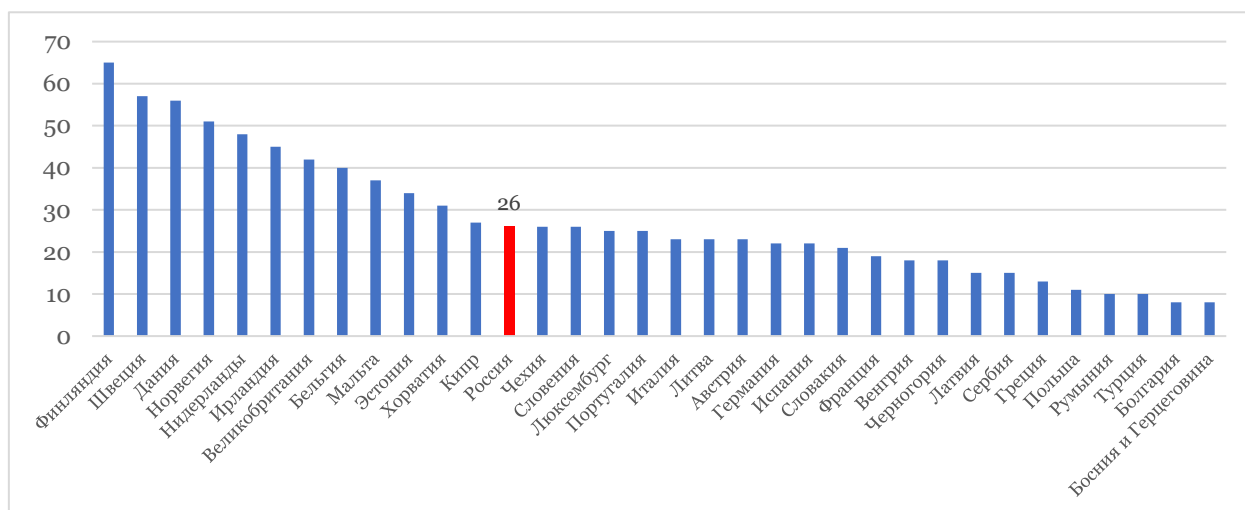


Рис. 12. Доля организаций, использующих сервисы облачных вычислений, %, 2018, ЕС, Россия
Источники: [43,44]

Отраслевой разрез по доле компаний использующих услуги облачных вычислений в 2019 году показывает, что лидерами среди различных сфер деятельности являются высшее образование (49%), финансовый сектор (38%) и здравоохранение (38%), их значения в том числе превышают и средний уровень для ЕС (30%). Использование услуг облачных вычислений в науке (29%) и промышленности (28%) соответствуют среднему значению по России, тогда как система государственного управления, транспорт, строительство и сельское хозяйство уступают средним общероссийским значениям, но незначительно (см. рисунок 13).

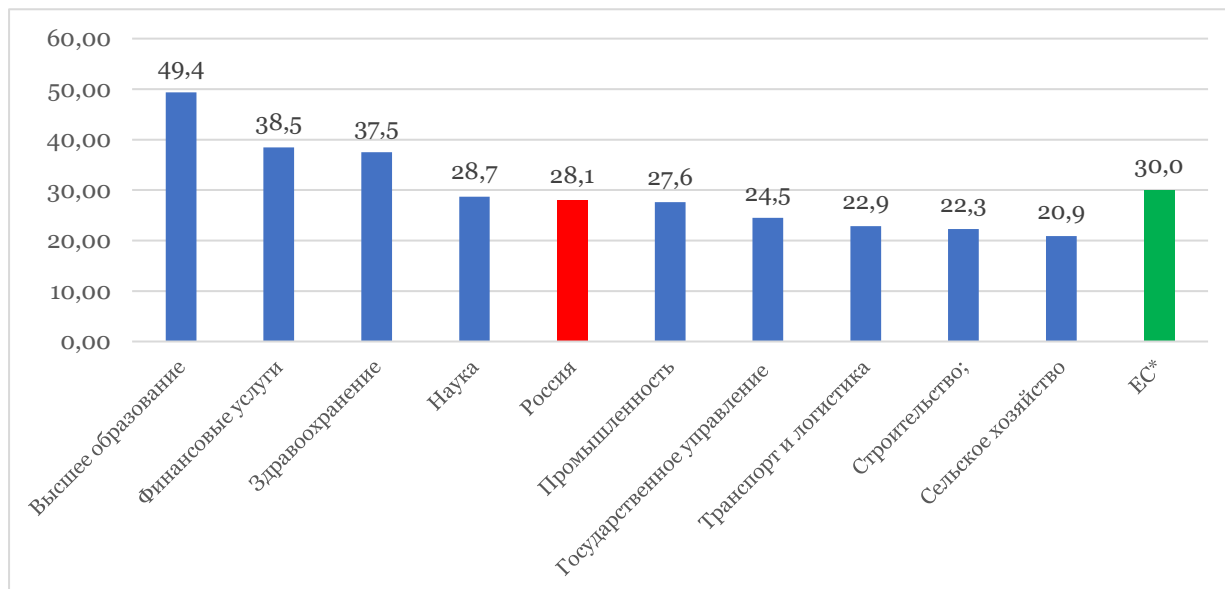


Рис. 13. Доля организаций, использующих сервисы облачных вычислений, %, в разрезе сфер деятельности, Россия (2019), ЕС (2018 – данные в 2019 году не собирались)
Источники: [43] [44]

4.3 Платформенная инфраструктура для работы с большими данными

Мониторинг платформенной инфраструктуры по предложенной системе показателей в настоящее время в рамках статистического наблюдения только зарождается: первые показатели использования цифровых платформ в 2020 году появились в обновленной форме №3-информ [48], хотя этих сведений для полноценной оценки наличия цифровых платформ различных типов для работы с большими данными явно недостаточно.

Поэтому в рамках системы мониторинга BD4DE предполагается, что данные для расчета показателей (ПИБД-03-01) – (ПИБД-03-04) будут собираться путем экспертных опросов с использованием специального инструментария (экспертной анкеты). Разработанная анкета прошла две стадии пилотирования.

На первой стадии набор показателей по мониторингу платформенной инфраструктуры прошел экспертную апробацию у специалистов по мониторингу развития цифровой экономики и статистики информационного общества. Ими были сформулированы рекомендации по расширению перечня показателей для учета дополнительных факторов, влияющих на применение технологий работы с большими данными. В частности, был предложен дополнительный показатель ПИБД-03-04 «Наличие экосистемы цифровых платформ», а по показателю ПИБД-03-02 «Наличие доступных специализированных цифровых платформ» было предложено собирать данные в разрезе отдельных сфер деятельности.

На второй стадии пилотирования системы мониторинга проводилась апробация анкеты экспертом в данной предметной области (специалистом по применению цифровых технологий в организациях предпринимательского сектора) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций), от которых были получены предложения по оптимизации структуры анкеты, снабжении анкеты необходимым понятийным аппаратом и уточнению формулировок отдельных вопросов.

По итогам проведения пилотирования экспертная анкета была доработана и подготовлена для проведения опроса, для повышения эффективности был разработан онлайн-вариант анкеты.

Заключение

Описанная концептуальная схема и набор показателей мониторинга и оценки цифровой инфраструктуры для работы с большими данными охватывает все уровни: от телекоммуникационной инфраструктуры передачи данных и вычислительной инфраструктуры для их хранения и обработки до платформенной инфраструктуры, предоставляющей возможность использовать результаты оперирования с большими данными в различных сферах деятельности. Некоторые показатели можно рассчитать на основе официальной международной статистики (МСЭ, ОЭСР, Евростат) и результатов федерального статистического наблюдения, регулярно публикуемого Росстатом, однако часть предложенных показателей может быть рассчитана только в результате проведения кабинетных исследований или экспертных опросов.

Важно отметить, что проведение кабинетных исследований для получения необходимых данных может быть затруднено в силу нерегулярного выпуска аналитических материалов, авторами которых часто выступают консалтинговые компании или независимые исследовательские организации, у которых нет публичных обязательств передавать соответствующие исследования в открытый доступ, поэтому часть необходимых данных может либо отсутствовать вовсе за нужный период времени, либо распространяться на коммерческой основе.

В последнее время система статистических наблюдений за цифровой инфраструктурой как со стороны международных организаций разного уровня (МСЭ, ОЭСР, Евростат), так и со стороны России активно модифицируется, в том числе идет активное включение в нее показателей, связанных с работой с большими данными. Дальнейшее развитие системы мониторинга BD4DE, связанное с цифровой инфраструктурой для работы с большими данными, должно учитывать изменения в системе статистических наблюдений международных и российских организаций и оперативного расширения перечня релевантных показателей.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. TeleGeography's Global Bandwidth Research Service. URL: <https://www2.telegeography.com/global-bandwidth-research-service> (дата обращения 01.06.2021)
2. Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (дата обращения 01.06.2021)
3. MarketsandMarkets. High Performance Computing (HPC) Market. URL: <https://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html> (дата обращения 01.06.2021)
4. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
5. Still, K., Seppänen, M., Korhonen, H., Valkokari, K., Suominen, A., & Kumpulainen, M. (2017). Business Model Innovation of Startups Developing Multisided Digital Platforms. In Y. Manolopoulos, M. Zaki, D. Diaz, & B. Theodoulidis (Eds.), 2017 IEEE 19th Conference on Business

- Informatics (CBI) (Vol. 2, pp. 70-75). [8012942] IEEE Institute of Electrical and Electronic Engineers. <https://doi.org/10.1109/CBI.2017.86>
6. Rossotto, C. M., Lal Das, P., Gasol Ramos, E., Clemente Miranda, E., Badran, M. F., Martinez Licetti, M., & Miralles Murciego, G. (2018). Digital platforms: A literature review and policy implications for development. *Competition and Regulation in Network Industries*, 19(1–2), 93–109. <https://doi.org/10.1177/1783591718809485>
 7. Yaroslav Eferin, Yuri Hohlov, Carlo Rossotto (2019) «Digital platforms in Russia: competition between national and foreign multi-sided platforms stimulates growth and innovation», *Digital Policy, Regulation and Governance*, Vol. 21, Issue 2, pp. 129-145. URL: <https://doi.org/10.1108/DPRG-11-2018-0065> (дата обращения 01.06.2021); Еферин Я.Ю., Россотто К.М., Хохлов Ю.Е. Цифровые платформы в России: конкуренция между национальными и зарубежными многосторонними платформами стимулирует экономический рост и инновации. *Информационное общество*, 2019, №1–2. С. 16–34. URL: <http://infosoc.iis.ru/article/view/135> (дата обращения 01.06.2021)
 8. Jacobides, MG, Cennamo, C, Gawer, A. Towards a theory of ecosystems. *Strat Mgmt J.* 2018; 39: 2255–2276. <https://doi.org/10.1002/smj.2904>
 9. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // *Информационное общество*. 2021. № 4–5. С. ??? <https://doi.org/10.1145/3453483.3454078>
 10. ITU. Measuring the Information Society Report. URL: <https://www.itu.int/pub/D-IND-ICTOI> (дата обращения 01.06.2021)
 11. World Telecommunication/ICT Indicators Database. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx> (дата обращения 01.06.2021)
 12. OECD Statistics by theme. URL: <https://stats.oecd.org/> (дата обращения 01.06.2021)
 13. Eurostat Statistics by theme. URL: <https://ec.europa.eu/eurostat/web/main/data/browse-statistics-by-theme> (дата обращения 01.06.2021)
 14. Федеральная служба государственной статистики. Официальная статистика. URL: <https://rosstat.gov.ru/folder/10705> (дата обращения 01.06.2021)
 15. OECD (2019), *An Introduction to Online Platforms and Their Role in the Digital Transformation*, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
 16. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. – 166 с. URL: <http://deca.iis.ru> (дата обращения 01.06.2021)
 17. Mossberger Karen, Tolbert Caroline J. Digital Citizenship and Digital Communities: How Technology Matters for Individuals and Communities. <https://doi.org/10.4018/IJEPR.20210701.0a2>
 18. Santos GL, Endo PT, Sadok D; Kelner J. When 5G Meets Deep Learning: A Systematic Review. URL: <https://doi.org/10.3390/a13090208>
 19. Seetharam Karthik, Kagiya Nobuyuki, Sengupta Partho P. Application of mobile health, telemedicine and artificial intelligence to echocardiography. <https://doi.org/10.1530/ERP-18-0081>
 20. Barricelli Barbara Rita; Casiraghi Elena; Fogli Daniela. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. <https://doi.org/10.1109/ACCESS.2019.2953499>
 21. Zacharov Igor, Arslanov Rinat, Gunin Maksim, Stefonishin Daniil, Bykov Andrey, Pavlov Sergey, Panarin Oleg, Maliutin Anton, Rykovanov Sergey, Fedorov Maxim. “Zhores” - Petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. <https://doi.org/10.1515/eng-2019-0059>
 22. Ved Mohit, Rizwanahmed B. Big Data Analytics in Telecommunication using state-of-the-art Big Data Framework in a Distributed Computing Environment: A Case Study. <https://doi.org/10.1109/COMPSAC.2019.00066>
 23. Zacharov I., Panarin O., Ryabinkin E., Izotov K., Teslyuk A. Virtualization for Scientific Workload. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000462194600017> (дата обращения 01.06.2021)
 24. Liao Jianwei, Gerofi Balazs, Lien Guo-Yuan, Miyoshi Takemasa, Nishizawa Seiya, Tomita Hirofumi, Liao Wei-Keng, Choudhary Alok, Ishikawa Yutaka. A flexible I/O arbitration framework for netCDF-based big data processing workflows on high-end supercomputers. URL:

- <https://www.webofscience.com/wos/woscc/full-record/WOS:000405351700005> (дата обращения 01.06.2021)
25. Casero-Ripolles Andreu. Influencers in the Political Conversation on Twitter: Identifying Digital Authority with Big Data. <https://doi.org/10.3390/su13052851>
 26. Sampugnaro Rossana, Montemagno Francesca. In Search of the Americanization: Candidates and Political Campaigns in European General Election. <https://doi.org/10.1080/15377857.2020.1869832>
 27. Al-Abdallah Ghaith, Khair Nadine, Elmarakby Reyad. The Impact of Social Networking Sites on Luxury Vehicles Purchase Decision Process in Gulf Cooperation Council Countries. <https://doi.org/10.1080/08961530.2020.1867023>
 28. Fiofanova Olga Aleksandrovna. New literacy and data-future in education: advanced technology smart big-data. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000572957200014> (дата обращения 01.06.2021)
 29. Yu Jun, Couldry Nick. Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. <https://doi.org/10.1080/1369118X.2020.1764604>
 30. Efimov V, Lapteva A. Digital technology in higher education: situation analysis and prospects assessment (on the example of Krasnoyarsk Krai). <https://doi.org/10.21125/inted.2019.1392>
 31. Larionova Viola, Sheka Andrey, Vasilyev Stanislaus. Students Behavioural Patterns on the National Open Education Platform. <https://doi.org/10.34190/EEL.19.126>
 32. Schiavone Francesco, Leone Daniele, Sorrentino Annarita, Scaletti Alessandro. Re-designing the service experience in the value co-creation process: an exploratory study of a healthcare network. <https://doi.org/10.1108/BPMJ-11-2019-0475>
 33. Pollock Daniel A. Digital Platforms as a Method of Invention for Infection Surveillance. <https://doi.org/10.1089/sur.2019.147>
 34. Aledavood Talayeh, Hoyos Ana Maria Triana, Alakorkko Tuomas, Kaski Kimmo, Saramaki Jari, Isometsa Erkki, Darst Richard K. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. <https://doi.org/10.2196/resprot.6919>
 35. Moller Arlen C., Merchant Gina, Conroy David E., West Robert, Hekler Eric, Kugler Kari C., Michie Susan. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. <https://doi.org/10.1007/s10865-016-9818-7>
 36. Ivaschenko Anton, Korchivoy Stanislav, Spodobaeв Michail. Infrastructural Models of Intermediary Service Providers in Digital Economy. https://doi.org/10.1007/978-3-030-29513-4_44
 37. Попов E.V. Econotronics of a smart city. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000476663600010> (дата обращения 01.06.2021)
 38. Kiraly, Peter; Buechler, Marco. Measuring completeness as metadata quality metric in Europeana. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000468499302103> (дата обращения 01.06.2021)
 39. Указ Президента Российской Федерации от 21.07.2020 г. № 424 «О национальных целях развития Российской Федерации до 2030 года». URL: <http://www.kremlin.ru/events/president/news/63728> (дата обращения 01.06.2021)
 40. ITU. Measuring digital development ICT price trends 2020. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceTrends_2020.pdf (дата обращения 01.06.2021)
 41. TOP 500. URL: <https://www.top500.org/lists/top500/> (дата обращения 15.06.2021)
 42. IDC. Локомотивом развития облачного рынка стали платформенные сервисы. URL: <https://www.idc.com/getdoc.jsp?containerId=prEUR247478021> (дата обращения 15.06.2021)
 43. Федеральная служба государственной статистики. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании информационных и коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказании услуг в этих сферах». URL: <https://rosstat.gov.ru/folder/14478> (дата обращения 01.10.2021)
 44. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (дата обращения 01.06.2021)
 45. ITU Statistics. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (дата обращения 01.06.2021)

46. Росстат. Итоги выборочного федерального статистического наблюдения по вопросам использования населением информационных технологий и информационно-телекоммуникационных сетей. URL: https://gks.ru/free_doc/new_site/business/it/ikt20/index.html (дата обращения 01.06.2021)
47. ITU. The historical data for the price baskets for countries and regions. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceBaskets_2008-2020.xlsx (дата обращения 01.06.2021)
48. Сведения об использовании информационных и коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказания услуг в этих сферах. № 3-информ. URL: <https://www.fedstat.ru/tools/storage/file?id=2106> (дата обращения 01.10.2021)

DIGITAL INFRASTRUCTURE FOR BIG DATA

Ershov, Peter Sergeevich

*Institute of the Information Society, head of Directorate of regional programs
Moscow, Russian Federation
peter.ershov@iis.ru*

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Abstract

A conceptual framework and a set of indicators for monitoring and evaluating digital infrastructure for big data have been developed. The conceptual framework consists of three components: telecommunication infrastructure; computing infrastructure; platform infrastructure. Calculations of the main indicators for Russia and individual fields of activity are given; international comparisons are made as of the end of 2020. It was noted that Russia lags significantly behind the EU countries in terms of certain infrastructural indicators.

Keywords

digital technology; big data; digital infrastructure; information infrastructure; telecommunication infrastructure; data transmission network; computing infrastructure; digital platform; monitoring and evaluation; Big Data for Digital Economy; BD4DE

References

1. TeleGeography's Global Bandwidth Research Service. URL: <https://www2.telegeography.com/global-bandwidth-research-service> (accessed 01.06.2021) Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (accessed 01.06.2021)
2. Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (accessed 01.06.2021)
3. MarketsandMarkets. High Performance Computing (HPC) Market. URL: <https://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html> (accessed 01.06.2021)
4. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
5. Still, K., Seppänen, M., Korhonen, H., Valkokari, K., Suominen, A., & Kumpulainen, M. (2017). Business Model Innovation of Startups Developing Multisided Digital Platforms. In Y. Manolopoulos, M. Zaki, D. Diaz, & B. Theodoulidis (Eds.), 2017 IEEE 19th Conference on Business Informatics (CBI) (Vol. 2, pp. 70-75). [8012942] IEEE Institute of Electrical and Electronic Engineers. <https://doi.org/10.1109/CBI.2017.86>
6. Rossotto, C. M., Lal Das, P., Gasol Ramos, E., Clemente Miranda, E., Badran, M. F., Martinez Licetti, M., & Miralles Murciego, G. (2018). Digital platforms: A literature review and policy implications for development. *Competition and Regulation in Network Industries*, 19(1–2), 93–109. <https://doi.org/10.1177/1783591718809485>
7. Yaroslav Eferin, Yuri Hohlov, Carlo Rossotto (2019) «Digital platforms in Russia: competition between national and foreign multi-sided platforms stimulates growth and innovation», *Digital Policy, Regulation and Governance*, Vol. 21, Issue 2, pp. 129-145. URL: <https://doi.org/10.1108/DPRG-11-2018-0065> (accessed 01.06.2021); Yeferin Ya.Yu, Rossotto K.M., Hohlov Yu.E. Tsifrovyye platformy v Rossii: konkurenciya mezhdu natsional'nymi i zarubezhnymi mnogostoronnimi platformami stimuliruyet ekonomicheskiy rost i

- innovatsii. Informatsionnoye obshchestvo, 2019, №1–2. С. 16–34. URL: <http://infosoc.iis.ru/article/view/135> (accessed 01.06.2021)
8. Jacobides, MG, Cennamo, C, Gawer, A. Towards a theory of ecosystems. *Strat Mgmt J.* 2018; 39: 2255–2276. <https://doi.org/10.1002/smj.2904>
 9. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // *Informatsionnoye obshchestvo.* 2021. № 4–5. С. ??? <https://doi.org/10.1145/3453483.3454078>
 10. ITU. Measuring the Information Society Report. URL: <https://www.itu.int/pub/D-IND-ICTOI> (accessed 01.06.2021)
 11. World Telecommunication/ICT Indicators Database. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx> (accessed 01.06.2021)
 12. OECD Statistics by theme. URL: <https://stats.oecd.org/> (accessed 01.06.2021)
 13. Eurostat Statistics by theme. URL: <https://ec.europa.eu/eurostat/web/main/data/browse-statistics-by-theme> (accessed 01.06.2021)
 14. Федеральная служба государственной статистики. Официальная статистика. URL: <https://rosstat.gov.ru/folder/10705> (accessed 01.06.2021)
 15. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
 16. Analiz tekushchego sostoyaniya razvitiya tsifrovoy ekonomiki v Rossii. М.: Institut razvitiya informatsionnogo obshchestva, 2018. – 166 s. URL: <http://deca.iis.ru> (accessed 01.06.2021)
 17. Mossberger Karen, Tolbert Caroline J. Digital Citizenship and Digital Communities: How Technology Matters for Individuals and Communities. <https://doi.org/10.4018/IJEPR.20210701.0a2>
 18. Santos GL, Endo PT, Sadok D; Kelner J. When 5G Meets Deep Learning: A Systematic Review. URL: <https://doi.org/10.3390/a13090208>
 19. Seetharam Karthik, Kagiya Nobuyuki, Sengupta Partho P. Application of mobile health, telemedicine and artificial intelligence to echocardiography. <https://doi.org/10.1530/ERP-18-0081>
 20. Barricelli Barbara Rita; Casiraghi Elena; Fogli Daniela. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. <https://doi.org/10.1109/ACCESS.2019.2953499>
 21. Zacharov Igor, Arslanov Rinat, Gunin Maksim, Stefonishin Daniil, Bykov Andrey, Pavlov Sergey, Panarin Oleg, Maliutin Anton, Rykovanov Sergey, Fedorov Maxim. “Zhores” - Petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. <https://doi.org/10.1515/eng-2019-0059>
 22. Ved Mohit, Rizwanahmed B. Big Data Analytics in Telecommunication using state-of-the-art Big Data Framework in a Distributed Computing Environment: A Case Study. <https://doi.org/10.1109/COMPSAC.2019.00066>
 23. Zacharov I, Panarin O., Ryabinkin E., Izotov K., Teslyuk A. Virtualization for Scientific Workload. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000462194600017> ((accessed 01.06.2021)
 24. Liao Jianwei, Gerofi Balazs, Lien Guo-Yuan, Miyoshi Takemasa, Nishizawa Seiya, Tomita Hirofumi, Liao Wei-Keng, Choudhary Alok, Ishikawa Yutaka. A flexible I/O arbitration framework for netCDF-based big data processing workflows on high-end supercomputers. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000405351700005> (accessed 01.06.2021)
 25. Casero-Ripolles Andreu. Influencers in the Political Conversation on Twitter: Identifying Digital Authority with Big Data. <https://doi.org/10.3390/su13052851>
 26. Sampugnaro Rossana, Montemagno Francesca. In Search of the Americanization: Candidates and Political Campaigns in European General Election. <https://doi.org/10.1080/15377857.2020.1869832>
 27. Al-Abdallah Ghaith, Khair Nadine, Elmarakby Reyad. The Impact of Social Networking Sites on Luxury Vehicles Purchase Decision Process in Gulf Cooperation Council Countries. <https://doi.org/10.1080/08961530.2020.1867023>
 28. Fiofanova Olga Aleksandrovna. New literacy and data-future in education: advanced technology smart big-data. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000572957200014> (accessed 01.06.2021)

29. Yu Jun, Couldry Nick. Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. <https://doi.org/10.1080/1369118X.2020.1764604>
30. Efimov V, Lapteva A. Digital technology in higher education: situation analysis and prospects assessment (on the example of Krasnoyarsk Krai). <https://doi.org/10.21125/inted.2019.1392>
31. Larionova Viola, Sheka Andrey, Vasilyev Stanislaus. Students Behavioural Patterns on the National Open Education Platform. <https://doi.org/10.34190/EEL.19.126>
32. Schiavone Francesco, Leone Daniele, Sorrentino Annarita, Scaletti Alessandro. Re-designing the service experience in the value co-creation process: an exploratory study of a healthcare network. <https://doi.org/10.1108/BPMJ-11-2019-0475>
33. Pollock Daniel A. Digital Platforms as a Method of Invention for Infection Surveillance. <https://doi.org/10.1089/sur.2019.147>
34. Aledavood Talayeh, Hoyos Ana Maria Triana, Alakorkko Tuomas, Kaski Kimmo, Saramaki Jari, Isometsa Erkki, Darst Richard K. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. <https://doi.org/10.2196/resprot.6919>
35. Moller Arlen C., Merchant Gina, Conroy David E., West Robert, Hekler Eric, Kugler Kari C., Michie Susan. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. <https://doi.org/10.1007/s10865-016-9818-7>
36. Ivaschenko Anton, Korchivoy Stanislav, Spodobaev Michail. Infrastructural Models of Intermediary Service Providers in Digital Economy. https://doi.org/10.1007/978-3-030-29513-4_44
37. Popov E.V. Econotronics of a smart city. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000476663600010> (accessed 01.06.2021)
38. Kiraly, Peter; Buechler, Marco. Measuring completeness as metadata quality metric in Europeana. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000468499302103> (accessed 01.06.2021)
39. Ukaz Prezidenta Rossiyskoy Federatsii ot 21.07.2020 g. № 424 «O natsional'nykh tselyakh razvitiya Rossiyskoy Federatsii do 2030 goda». URL: <http://www.kremlin.ru/events/president/news/63728> (accessed 01.06.2021)
40. ITU. Measuring digital development ICT price trends 2020. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceTrends_2020.pdf (accessed 01.06.2021)
41. TOP 500. URL: <https://www.top500.org/lists/top500/> (accessed 01.06.2021)
42. IDC. Lokomotivom razvitiya oblachnogo rynka stali platformnyye servisy. URL: <https://www.idc.com/getdoc.jsp?containerId=prEUR247478021> (accessed 01.06.2021)
43. Federal'naya sluzhba gosudarstvennoy statistiki. Itogi federal'nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol'zovanii informatsionnykh i kommunikatsionnykh tekhnologiy i proizvodstve vychislitel'noy tekhniki, programmnoy obespecheniya i okazanii uslug v etikh sferakh». URL: <https://rosstat.gov.ru/folder/14478> (accessed 01.06.2021)
44. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (accessed 01.06.2021)
45. ITU Statistics. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed 01.06.2021)
46. Rosstat. Itogi vyborochnogo federal'nogo statisticheskogo nablyudeniya po voprosam ispol'zovaniya naseleniyem informatsionnykh tekhnologiy i informatsionno-telekommunikatsionnykh setey. URL: https://gks.ru/free_doc/new_site/business/it/ikt20/index.html (accessed 01.06.2021)
47. ITU. The historical data for the price baskets for countries and regions. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceBaskets_2008-2020.xlsx (accessed 01.06.2021)
48. Svedeniya ob ispol'zovanii informatsionnykh i kommunikatsionnykh tekhnologiy i proizvodstve vychislitel'noy tekhniki, programmnoy obespecheniya i okazaniya uslug v etikh sferakh. № 3-inform. URL: <https://www.fedstat.ru/tools/storage/file?id=2106> (accessed 01.06.2021)

Социально-экономические аспекты информационного общества**БОЛЬШИЕ ДАННЫЕ: СОЦИАЛЬНЫЕ И ЭКОНОМИЧЕСКИЕ ЭФФЕКТЫ****Малахов Вадим Александрович**

Кандидат исторических наук

Институт истории естествознания и техники им. С. И. Вавилова РАН, Центр истории организации науки и науковедения, старший научный сотрудник

Москва, Российская Федерация

yasonbh@mail.ru

Хохлов Юрий Евгеньевич

Кандидат физико-математических наук, доцент

Институт развития информационного общества, председатель совета директоров

РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО

Москва, Российская Федерация

yuri.hohlov@iis.ru

Шапошник Сергей Борисович

Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития, старший научный сотрудник

Петрозаводск, Российская Федерация

sergei.shaposhnik@gmail.com

Юревич Максим Андреевич

Финансовый университет при Правительстве РФ, Центр макроэкономических исследований, научный сотрудник

Москва, Российская Федерация

mayurevich@fa.ru

Аннотация

Разработаны концептуальная схема и набор показателей для мониторинга экономических и неэкономических эффектов от использования технологий работы с большими данными. В концептуальной схеме два основных блока показателей: первый посвящен воздействию технологий работы с большими данными на экономику страны, второй – на другие сферы (общество, экология, безопасность). Апробация концептуальной схемы мониторинга BD4DE была проведена с помощью экспертного опроса, по результатам которого были рассчитаны коэффициенты косвенных эффектов; при этом значения этих показателей оказались сравнимы с показателями по странам Евросоюза и Великобритании. По результатам экспертного опроса были получены значения для показателей, характеризующих неэкономические (в том числе – социальные) эффекты от использования данного типа цифровых технологий. Выявлено, что наиболее существенное влияние работа с большими данными оказывает на улучшение качества предоставляемых услуг (как органами государственной власти, так и частным сектором), тогда как относительно влияния на рост или сокращение занятости населения мнения экспертов разделились.

Ключевые слова

большие данные; рынок больших данных, цифровая экономика, технологии работы с большими данными, мониторинг развития и использования цифровых технологий; Big Data for Digital Economy; BD4DE; социальный эффект; экономический эффект

© Малахов В.А., Хохлов Ю.Е., Шапошник С.Б., Юревич М.А., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_132

Введение

Благодаря развитию цифровых технологий в современном мире генерируются и доступны для анализа невиданные ранее массивы данных. Беспрецедентный рост объемов данных и возможностей по их сохранению и обработке оказывают существенное влияние на мировую экономику, социальную сферу, политику и культуру [1]. Сегодня сложно найти сферу человеческой деятельности, в которой не применялись бы большие данные: компании используют технологии обработки больших данных, чтобы выявить закономерности и максимизировать прибыль [2]; в здравоохранении подробные данные о состоянии здоровья населения помогают при профилактике эпидемий, излечении болезней и снижении затрат [3] – и это далеко не единственные примеры. Однако не все эффекты от использования больших данных являются однозначно положительными. Так, исследователи выражали озабоченность в связи с использованием больших данных для создания инвазивного маркетинга [4] или вторжения в личную жизнь граждан, в том числе со стороны власти [5].

Хотя влияние использования больших данных на экономику и социальную сферу очевидно, для принятия взвешенных решений как в бизнесе, так и в государственном секторе необходимо достаточно точно оценивать и измерять ожидаемое воздействие, как позитивное, так и негативное [6]. Например, правительству важно понимать, насколько значимыми будут экономические эффекты от внедрения технологий работы с большими данными в различных отраслях экономики, а также каковы будут последствия такого внедрения для занятости населения.

Один из подходов к мониторингу и оценке производства, использования и воздействия технологий, решений и услуг для работы с большими данными описан в работе [7] и развивается в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» Национальной технологической инициативы.

Данное исследование направлено на разработку одного из ключевых компонентов концептуальной схемы мониторинга BD4DE (Big Data for Digital Economy), связанного с социальными и экономическими эффектами. Предметной областью мониторинга являются как экономические эффекты, измеряемые количественно (например, объем рынка больших данных), так и неэкономические эффекты, которые не всегда возможно измерить количественно (например, влияние использования технологий работы с большими данными на экологическую ситуацию в стране и в мире).

Статья состоит из трех разделов: первый посвящен обзору подходов к мониторингу экономических и социальных эффектов использования технологий работы с большими данными в мире, второй – описанию разработанной концептуальной схемы мониторинга социальных и экономических эффектов от использования технологий работы с большими данными, третий – эконометрическому анализу воздействия технологий работы с большими данными на социально-экономические макропараметры. Разработанная концептуальная схема может быть применена для мониторинга социальных и экономических эффектов от использования технологий работы с большими данными как в России, так и в других странах.

1 Обзор подходов к мониторингу экономических и социальных эффектов использования технологий работы с большими данными

1.1 Показатели воздействия цифровых технологий в международных стандартах статистического наблюдения за использованием цифровых технологий

Как отмечено в [7], первоначальные методологии мониторинга и оценки использования ИКТ, так называемые методологии оценки электронной готовности (e-readiness assessment), не включали показатели, характеризующие социально-экономические эффекты использования технологий.

Разработка подходов и стандартов статистического наблюдения за воздействием технологий началось в начале 2000-х годов. Основной разработчик стандартов измерения производства и использования цифровых технологий ОЭСР в одном из первых вариантов модельной анкеты для бизнеса по использованию ИКТ, опубликованном в 2005 году, включила вопрос про выгоды, которые компания получает от использования электронных продаж (первый вариант модельной анкеты ОЭСР концентрировался в основном на электронной коммерции) [8]. Для респондентов предлагалось выбрать следующие варианты ответов:

- 1) сокращенное время транзакции;
- 2) повышение качества обслуживания клиентов;
- 3) снижение затрат, включая транзакционные и другие расходы;
- 4) увеличение объема продаж и/или количества клиентов;
- 5) сохранение конкурентных позиций;
- 6) возможность лучше учитывать индивидуальные особенности клиентов.

Такой подход в руководстве ОЭСР по измерению информационного общества [8] был назван «субъективным измерением воздействий». Суть этого подхода к измерению воздействия ИКТ на деятельность компаний заключается в том, чтобы напрямую спрашивать компании об этом воздействии. Как отмечается в руководстве, хотя этот подход дает преимущество в предоставлении прямой причинно-следственной информации, он обычно считается менее объективным, чем эмпирические методы измерения (на уровне предприятий, отраслей и экономики в целом).

Этот подход получил дальнейшее развитие в стандартах ОЭСР и Евростата. В модельную анкету Евростата для бизнеса по использованию ИКТ в 2008 г. был включен модуль о выгодах использования ИКТ [9]. Вопрос формулировался следующим образом. «В какой степени реализованные за последние 2 года проекты в области ИКТ привели к улучшениям в следующих областях по сравнению с предыдущим выполнением задач (до внедрения ИКТ):

- a) реорганизация и упрощение рабочих процессов;
- b) экономия ресурсов;
- c) увеличение выручки предприятия;
- d) разработка новых продуктов и услуг.

При этом для каждой области необходимо было указать степень воздействия: (1) незначительное/отсутствует, (2) умеренное, (3) существенное, (4) неизвестно/неприменимо.

В последнем варианте модельного обследования использования ИКТ бизнесом, опубликованном ОЭСР в 2015 г. [10], показатели «субъективного измерения воздействия» представлены в двух видах.

Первый вариант – включение вопроса об эффектах в модуль вопросов, связанных с конкретной технологией.

Так в модуле про использование облачных вычислений предлагается выбрать следующие варианты полученных выгод:

- a) сокращение затрат, связанных с ИКТ;
- b) гибкость в масштабировании сервисов;
- c) простота (легкость и быстрота) развертывания облачных решений;
- d) повышение производительности.

Измерение предлагается проводить двумя методами (на выбор): либо просто рассчитывать долю предприятий, получивших соответствующий эффект, либо проводить самооценку с использованием качественной шкалы – указывать для каждого эффекта степень его выраженности (например, высокая / небольшая / ограниченная / нет эффекта).

Аналогично в модуле вопросов про использование аналитики больших данных, предлагается оценить воздействие аналитики данных со следующими вариантами эффектов:

- a) экономия средств;
- b) рост продаж (в том числе за счет улучшения продукции и более эффективного маркетинга);
- c) улучшения в организации бизнеса.

Самооценка может выполняться в шкале «да / нет» или на основе качественной шкалы (например, [воздействие] высокое / частичное / ограниченное / отсутствует) для каждого вида воздействия.

Второй вариант использования показателей воздействия в модельном обследовании ОЭСР – выделение отдельного модуля показателей («Внедрение ключевых инструментов ИКТ: воспринимаемые преимущества, препятствия и воздействия – открытые индикаторы»), который относится к общим (базовым) технологиям и включает два вопроса [10]:

(1) Выгоды от выбранной ИКТ, включая следующие их варианты (варианты воздействия могут быть с градацией – сильное, ограниченное, отсутствует / неприменимо):

- a) снижение операционных затрат и/или затрат на рабочую силу;
- b) повышение способности реагировать на требования клиентов или поставщиков;
- c) сохранение конкурентных позиций;
- d) сокращение времени транзакции;
- e) повышение качества товаров или услуг;
- f) повышение гибкости производства или предоставления услуг;
- g) улучшение обмена информацией.

(2) Изменения в отдельных аспектах организации бизнеса в результате внедрения ИКТ (вопрос может применяться либо к конкретным приложениям или в более общих терминах):

- a) реинжиниринг бизнес-процессов;
- b) сбор данных, хранение и обслуживание;
- c) большая опора на должностную ротацию, мультиквалификацию.

Что касается эмпирических методов измерения эффектов, ОЭСР с начала 2000-х годов выпустила несколько докладов по расчету макроэкономических эффектов инвестиций и использования ИКТ (с использованием эконометрических методов) и в руководстве по измерению информационного общества приводит обзор таких макроэкономических работ (см. раздел 1.3, где приведен обзор новых исследований с применением подобного подхода для измерения эффектов от внедрения технологий работы с большими данными). В руководстве также отмечается важность эмпирического измерения воздействия ИКТ на уровне предприятий и анализируются проблемы с данными для такого рода исследований, конкретных рекомендаций и стандартов такого измерения не приводится [8].

В недавнем докладе ОЭСР «Измерение цифровой трансформации: дорожная карта для будущего» [11] два из девяти направлений связаны с измерением экономических и социальных эффектов. В разделе «Понимание экономических последствий цифровой трансформации» отмечается, что первоначальные и убедительные доказательства экономического воздействия скорее всего появятся в микроданных (данных о фирмах, работниках или потребителях), прежде чем проявиться в макроданных; даются общие рекомендации по проведению эмпирических исследований в этом направлении, в том числе с использованием административных данных. В частности, рекомендуется использовать возможности агрегирования на уровне предприятий статистических данных о цифровой зрелости и широкого спектра показателей экономической деятельности, полученных из различных форм статистического наблюдения, что позволит проводить статистические исследования экономических эффектов, в том числе лонгитюдные. Агрегирование широкого спектра показателей (в частности показателей использования ИКТ и различных аспектов жизнедеятельности) и лонгитюдные исследования рекомендуется также применять для измерения воздействия цифровой трансформации на социальные цели и благосостояние людей. Это позволит выявлять искомые причинно-следственные связи, в том числе между распространением цифровых технологий и различными результатами в плане благополучия. Подобные рекомендации, в том числе применительно к эффектам от работы с большими данными, вполне могут быть реализованы в России, но для этого требуется изменение практики работы Росстата с итогами различных обследований.

В той или иной степени описанные выше подходы, включая «субъективное измерение воздействий», реализованы в рамках данной системы мониторинга BD4DE.

1.2 Социологические исследования и анализ статистики национальных счетов

Сегодня над созданием инструментария для измерения цифровой экономики и эффектов использования технологий для работы с большими данными работают многие международные и национальные организации. Так, в рамках Группы двадцати начиная с 2016 года функционирует экспертная группа по цифровой экономике, занимающаяся подготовкой документов с рекомендациями по стандартизации мониторинга и оценки цифровой экономики [12].

Как правило, зарубежная практика мониторинга величины цифровой экономики и, в частности, экономики больших данных строится на экспертном отборе подходящих видов экономической деятельности и подсчета совокупной добавленной стоимости по отобранным видам за соответствующий временной интервал. Такой подход для цифровой экономики был впервые разработан и успешно применен Бюро экономического анализа США [13], а впоследствии скопирован Статистической службой Канады [14] и Австралийским бюро статистики [15].

В Великобритании оригинальное исследование по оценке объема экономики больших данных и интернета вещей было проведено Центром деловых и экономических исследований (CEBR) [16]. Методология CEBR опирается на анализ результатов опросов представителей бизнеса. По результатам опроса для всех отраслей экономики Великобритании и для системы государственного управления рассчитывается уровень внедрения технологий работы с большими данными и интернета вещей (в процентах), а также средний уровень положительных эффектов от внедрения данных технологий в отдельных сферах деятельности. На основе данных показателей производится оценка объема экономики больших данных и интернета вещей по отраслям и в целом по экономике. Особенностью использованного подхода является разделение экономических эффектов использования технологий работы с большими данными на три блока:

- воздействие на эффективность предприятий (улучшение производительности, снижение затрат и т.д.);
- воздействие на инновации в бизнесе (инновации в бизнесе, возникшие в результате увеличения нераспределенной прибыли для инвестиций в инновации и НИОКР);
- воздействие, приводящее к созданию нового бизнеса (экономическая активность, генерируемая во всей экономике в результате повышения эффективности бизнеса и появления дополнительных денежных потоков).

Помимо измерения объема экономики больших данных в отчете CEBR даны экспертные оценки влияния использования технологий работы с большими данными на уровень занятости населения в Великобритании.

Для измерения экономических и социальных эффектов от внедрения и использования технологий, решений и услуг для работы с большими данными в рамках исследования Европейского рынка данных по заказу Европейской комиссии был разработан более сложный алгоритм, сочетающий в себе как данные Евростата, так и результаты опроса экспертов и предприятий [17]. На первом этапе исследований с помощью официальной статистики Евростата был определен объем рынка больших данных. Для этого в классификаторе экономических отраслей Евростата (NACE rev. 2) [18] экспертами были отобраны отрасли, в которых работают поставщики (производители) технологий, решений и услуг для работы с большими данными. В общей сложности было отобрано 11 связанных с ИКТ отраслей. Доля компаний, являющихся поставщиками, от всех компаний в данных отраслях была определена экспертным путем. Затем, исходя из доли компаний, являющихся поставщиками, и статистических данных об объеме рынка по каждой из отраслей был оценен объем рынка больших данных (прямые экономические эффекты) по странам ЕС.

Следующим шагом после расчета объема рынка больших данных стал расчет косвенных экономических эффектов от внедрения технологий, решений и услуг для работы с большими данными, которые включают:

1. Обратные косвенные эффекты (англ. backward indirect impacts) – формируются в результате спроса поставщиков на оборудование, человеческий капитал и сопутствующие товары и услуги, которые им необходимы для производства и предоставления информационной продукции.
2. Форвардные косвенные эффекты (англ. forward indirect impacts) – возникают в результате внедрения в экономические процессы продукции поставщиков, т.е. объем экономической выгоды, возникающей у организаций, использующих технологии, решения и услуги для работы с большими данными (далее – потребители). Например, оптимизация процессов производства, логистики, маркетинга, повышение эффективности менеджмента.
3. Индуцированные косвенные эффекты (англ. induced impacts) – включают в себя экономическую активность, генерируемую во всей экономике (вторичный эффект). Дополнительные денежные потоки возникают как у поставщиков, так и у потребителей (за счет увеличения прибыли, роста фонда заработной платы и т.д.). Эти дополнительные потоки порождают, в свою очередь, дополнительное потребление, которое поддерживает экономическую активность в различных отраслях, таких как розничная торговля, потребительские товары, банковский сектор, развлечения и т. д. Эти эффекты отражаются почти во всех секторах экономики.

Все косвенные экономические эффекты рассчитывались экспертным путем в ходе экспертного опроса (экспертам предлагалось оценить коэффициент объема каждого косвенного экономического эффекта по отношению к прямому, например: 1 евро прямых экономических

эффектов генерирует 0,07 евро обратных косвенных эффектов). Помимо мониторинга экономических эффектов от внедрения технологий работы с большими данными в докладе [17] даны оценки некоторых социальных эффектов, таких как влияние больших данных на рынок труда в ЕС (на основе статистики Евростата и экспертных оценок).

1.3 Эконометрические подходы

Для анализа релевантной научной литературы по исследуемой теме были применены традиционные методы библиометрического поиска и анализа. Формулировка поискового запроса “big data” AND (societal OR economic*) AND (effects OR impact)” при поиске по всем доступным полям позволила обнаружить более 2 тыс. публикаций в библиографической базе данных платформы Web of Science (WoS). Выбор дисциплинарных областей “Economics” и “Management” сузил первоначальный массив публикаций до 404 единиц. Далее в приоритетном порядке рассматривались работы, имеющие высокий уровень цитирований, а также статьи, находящиеся в списках литературы или цитирующие наиболее подходящие исследования. Дополнительный поиск был выполнен в поисковой системе Google и библиографической базе данных Google Scholar, что позволило выявить дополнительные аналитические отчеты и публикации, не имеющие индексацию в WoS. В итоге было отобрано 10 наиболее релевантных научных статей и отчетов [19-29], посвященных использованию количественных методов для оценки экономических и социальных эффектов использования технологий работы с большими данными.

Эконометрические модели оценки влияния технологий больших данных на индикаторы экономической эффективности преимущественно строятся по данным деятельности отдельных организаций, так как расчеты эффектов на уровне отдельных отраслей или национальных экономик сильно затруднены из-за ограниченности исходной статистической информации [19]. Исследования на микроуровне опираются на модели непосредственной регрессионной зависимости, когда в организации наблюдается прямой эффект от использования больших данных на производительность труда, размер выручки или прибыли, а также на алгоритмы оценки опосредованной зависимости, когда применение больших данных повышает уровень инновационной активности, а та в свою очередь стимулирует рост экономической эффективности (например, через модели Крепона-Дугата-Морисса [30]).

Опрос высшего менеджмента (технических директоров, директоров по информационным технологиям) организаций из различных отраслей позволил выявить положительное воздействие внедрения технологий анализа больших данных на рыночные и операционные характеристики организаций, в том числе получение конкурентных преимуществ [20]. Аналогичный результат был достигнут при моделировании зависимости между использованием в производственной деятельности аналитики больших данных и набором индикаторов финансовой успешности организаций (объем продаж, прибыльность, окупаемость инвестиций и т.д.) [21].

Использование больших данных управленческим персоналом также воздействует на показатели деятельности предприятий. Так, среди публичных американских компаний практика принятия управленческих решений на основе больших данных генерировала прибавку в 5–6% к производительности и объему продаж [22]. В другом исследовании удалось зафиксировать эффект в 3% роста производительности труда от такого стиля менеджмента среди предприятий обрабатывающей промышленности [23]. Наем персонала, обладающего навыками работы с большими данными, включая их сбор и аналитику, стимулирует рост продаж, производительности труда и инновационной активности организаций [24]. При этом сравнительно больший эффект наблюдается в организациях с децентрализованным управлением. Трудоустройство сотрудников с навыками работы в среде Nadoor форсировало рост производительности труда в организациях, накопивших значительные объемы данных о своих бизнес-процессах; привлечение специалистов в области традиционных технологий обработки данных, таких как реляционные базы данных, ощутимой выгоды не дало [25].

Положительное воздействие применения аналитики больших данных было обнаружено в отношении склонности организаций внедрять как продуктовые, так и процессные инновации. Однако в сфере услуг эта связь оказалась более устойчивой по сравнению с производственным сектором [26]. В одном из российских исследований было показано, что анализ больших объемов данных и применение предсказательной аналитики в значительно большей степени свойственны компаниям-лидерам отраслевых отечественных рынков [27].

Интересно, что широкое распространение аналитики данных продуцирует наибольшую выгоду, когда организация главным образом ориентирована на улучшающие инновации, а не на создание принципиально новых технологий. Эти результаты согласуются с теорией о том, что аналитика данных дополняет определенные виды инноваций, поскольку позволяет организации расширить пространство поиска существующих знаний для объединения в новые технологии, а также с теоретическими аргументами о том, что аналитика данных поддерживает постепенное совершенствование технологических процессов [28].

В одном из немногих макроэкономических исследований было установлено влияние интенсивности использования технологий работы с большими данными на рост ВВП в Великобритании [29]. В основу модели положена система уравнений, отражающих этапы технологического трансфера (по аналогии с моделью Крепона-Дугата-Морисса), а исходные данные для расчетов были получены от Национальной статистической службы. Переменная, отражающая использование больших данных, была сконструирована по алгоритму, близкому к использовавшемуся в уже упомянутом выше проекте Еврокомиссии [17]. В итоге авторы исследования получили следующий результат: в 2012 г. индустрия больших данных Великобритании сгенерировала около 150 млн фунтов стерлингов дополнительной прибавочной стоимости или обеспечила 0,02% из 1,49% роста ВВП.

2 Концептуальная схема и набор показателей мониторинга

2.1 Концептуальная схема предметной области

Результаты анализа из предыдущих разделов приводят к выделению в концептуальной схеме мониторинга и оценки воздействия технологий работы с большими данными на социально-экономическое развитие двух основных компонентов: (1) экономические эффекты; (2) неэкономические эффекты. Кроме того, в концептуальной схеме должна быть реализована возможность измерения воздействия от использования технологий работы с большими данными как на макроуровне (экономику и общество в целом), так и на уровне отдельных организаций или отраслей.

Данный подход позволит не только оценить долговременные последствия внедрения новых технологий на экономику, общество и окружающую среду в целом, но и поможет определить целесообразность их внедрения на уровне конкретных организаций или в разрезе отдельных отраслей (см. таблицу 1).

Таблица 1. Концептуальная схема мониторинга эффектов использования технологий работы с большими данными

Экономические эффекты от использования технологий работы с большими данными	Неэкономические эффекты от использования технологий работы с большими данными
Прямые экономические эффекты	Влияние на безопасность жизнедеятельности в бытовой и профессиональной областях
Обратные косвенные эффекты	Влияние на повышение качества предоставляемых услуг
Форвардные косвенные эффекты	Влияние на экологическую ситуацию в стране
Индукцированные косвенные эффекты	Влияние технологий работы с большими данными на занятость населения
Доля организаций, увеличивших выручку за счет использования технологий работы с большими данными	Доля организаций, повысивших качество товаров и услуг за счет использования технологий работы с большими данными
Доля организаций, снизивших издержки за счет использования технологий работы с большими данными	Доля организаций, снизивших экологический ущерб за счет использования технологий работы с большими данными
Доля организаций, повысивших производительность труда за счет использования технологий работы с	Доля организаций, улучшивших работу с кадрами за счет использования технологий работы с большими данными

Экономические эффекты от использования технологий работы с большими данными	Неэкономические эффекты от использования технологий работы с большими данными
большими данными	
	Доля организаций, улучшивших взаимодействие с контрагентами за счет использования технологий работы с большими данными
	Доля организаций, повысивших адаптивность/гибкость процессов за счет использования технологий работы с большими данными
	Доля организаций, повысивших безопасность труда сотрудников за счет использования технологий работы с большими данными
	Доля организаций, повысивших качества принимаемых управленческих решений

Выбор перечня макроэкономических показателей системы мониторинга BD4DE в значительной степени опирался на методологию, использованную в проекте Европейской комиссии [17] с целью проведения дальнейших межстрановых сопоставлений ситуации в России и странах Евросоюза.

Измерение макроэкономических эффектов от внедрения и использования технологий больших данных соответствует аналогичному разделу мониторинга экономики больших данных стран Евросоюза и включает в себя показатели, характеризующие прямые и различные виды косвенных экономических эффектов.

Помимо измерения экономических эффектов от внедрения и использования новых технологий важно оценить их воздействие и на неэкономические сферы – экологию, общество, безопасность, изменение качества товаров и услуг.

Детальное описание предлагаемых показателей и их направленности дано в следующих разделах 2.2 и 2.3.

Апробация разработанной концептуальной схемы мониторинга может быть осуществлена с помощью опроса экспертов и организаций-поставщиков технологий, решений и услуг для работы с большими данными, а также организаций, использующих данные технологии, решения и услуги.

2.2 Показатели экономических эффектов

Мониторинг макроэкономических эффектов от использования технологий для работы с большими данными осуществляется по 4 показателям:

1. Прямые экономические эффекты – выручка производителей (поставщиков) технологий, решений, товаров и услуг для работы с большими данными.
2. Обратные косвенные эффекты – формируются в результате спроса поставщиков на оборудование, человеческий капитал и другие сопутствующие товары и услуги, которые им необходимы для производства и предоставления информационной продукции.
3. Форвардные косвенные эффекты – возникают в результате внедрения в экономические процессы продукции поставщиков, т.е. объем экономической выгоды, возникающей у потребителей.
4. Индуцированные косвенные эффекты – включают в себя экономическую активность, генерируемую во всей экономике (вторичный эффект).

Раздел, посвященный экономическим эффектам от внедрения новых технологий на деятельность организаций состоит из трех показателей: (1) доля организаций, увеличивших выручку за счет использования технологий работы с большими данными; (2) доля организаций, снизивших издержки за счет использования технологий работы с большими данными и (3) доля организаций, повысивших производительность труда за счет использования технологий работы с

большими данными. Это ключевые показатели, по которым можно судить об экономической целесообразности внедрения новых технологий в конкретных организациях. Стоит отметить, что данные показатели также могут применяться для измерения эффектов от использования технологий работы с большими данными в масштабах страны или отдельных экономических отраслей. Основное отличие данного блока показателей от других показателей экономического воздействия заключается в методе сбора данных. Данные показатели рассчитываются без привлечения экспертов, источником данных для них является опрос организаций.

2.3 Показатели неэкономических эффектов

Для оценки неэкономических эффектов от использования технологий для работы с большими данными в системе мониторинга BD4DE был разработан блок из четырех показателей:

1. Влияние на безопасность жизнедеятельности в бытовой и профессиональной областях. Повышение безопасности в самых различных сферах жизнедеятельности является одной из важнейших характеристик, по которым можно оценить неэкономические эффекты от использования технологий для работы с большими данными. Так, например, новые технологии во многом преобразили рынок средств для защиты компьютеров. Измерение показателя производится с помощью экспертного опроса по пятибалльной шкале, где 1 балл означает отсутствие влияния, а 5 баллов – максимально сильное влияние новых технологий в данной сфере.
2. Влияние на повышение качества предоставляемых услуг (как в государственном, так и в частном секторах). Анализ больших данных помогает выявлять тенденции развития рынка, предпочтения потенциальных клиентов и тем самым повышает уровень предоставляемых услуг. Измерение показателя производится с помощью экспертного опроса по пятибалльной шкале, где 1 балл означает отсутствие влияния, а 5 баллов – максимально сильное влияние новых технологий в данной сфере.
3. Влияние на экологическую ситуацию в стране. Данный показатель может принимать как положительное, так и отрицательное значение. Так, негативными эффектами от развития технологий работы с большими данными может стать повышение уровня потребления электроэнергии (характерно для некоторых видов технологий распределенных реестров), с другой стороны, применение больших данных может способствовать оптимизации эффективности в энергетическом секторе. Измерение показателя производится с помощью экспертного опроса по пятибалльной шкале, где 1 балл означает максимально негативное влияние, 3 балла – отсутствие влияния, 5 баллов – максимальное положительное влияние новых технологий в данной сфере.
4. Влияние на занятость населения. Данный показатель может принимать как положительное, так и отрицательное значение. Так, использование технологий работы с большими данными может способствовать как росту занятости населения (появление новых вакансий, связанных с разработкой и внедрением технологий), так и его сокращению за счет оптимизации труда. Измерение показателя производится с помощью экспертного опроса по пятибалльной шкале, где 1 балл означает максимально негативное влияние, 3 балла – отсутствие влияния, 5 баллов – максимальное положительное влияние новых технологий в данной сфере.

Раздел по неэкономическим эффектам от использования технологий работы с большими данными в организации включает следующие показатели: (1) доля организаций, повысивших качество товаров и услуг за счет использования технологий работы с большими данными; (2) доля организаций, снизивших экологический ущерб за счет использования технологий работы с большими данными; (3) доля организаций, улучшивших работу с кадрами за счет использования технологий работы с большими данными; (4) доля организаций, улучшивших взаимодействие с контрагентами за счет использования технологий работы с большими данными; (5) доля организаций, повысивших адаптивность/гибкость процессов за счет использования технологий работы с большими данными; (6) доля организаций, повысивших безопасность труда сотрудников за счет использования технологий работы с большими данными; (7) доля организаций, повысивших качество принимаемых управленческих решений. Как и в случае с показателями экономического воздействия данный блок показателей отличается от других показателей неэкономических эффектов методом сбора данных (опрос организаций); эти

показатели могут использоваться для измерения воздействия от использования технологий работы с большими данными в масштабах всей страны или отдельных сфер деятельности.

3 Регрессионный анализ воздействия больших данных на социально-экономические макропараметры

Мониторинг использования больших данных на макроэкономическом уровне пока не стал частью процесса регулярных статистических наблюдений международных и национальных организаций. Исключение составляет опрос Евростата [31], по результатам которого оценивается доля организаций, использующих большие данные из любых источников, в общем числе предприятий (переменная BD). Было проведено две волны опроса в 2016 и 2018 гг. в 29 странах.

Для проверки гипотез о воздействии больших данных на социально-экономические макропараметры используются следующие зависимые переменные, отражающие как размер экономики, уровень ее технологичности и развитости, степень цифровизации общества и экономики, так и совокупное благополучие граждан (через Индекс счастья):

GDP – ВВП в постоянных ценах по ППС, млн евро (источник: Евростат);

LP – производительность труда на один отработанный час, в % к среднему уровню по ЕС (источник: Евростат);

GDPpc – ВВП на душу населения в постоянных ценах по ППС, млн евро (источник: Евростат);

EXP – объем высокотехнологичного экспорта в постоянных ценах по ППС, млн евро (источник: Евростат);

EXPs – доля высокотехнологичного экспорта в общем объеме экспорта, % (источник: Евростат);

ECOMs – доля доходов от электронной торговли в валовой выручке предприятий, % (источник: Евростат);

HNAP – Индекс счастья, ед. (источник: World Happiness Report 2020 [32]);

DESI – Индекс цифровой экономики и общества, ед. (источник: Digital Economy and Society Index [33]);

GCI – Глобальный индекс коннективности, ед. (источник: Global Connectivity Index [34]).

Очевидно, что большие данные едва ли могут претендовать на роль ключевого драйвера экономического и социального благополучия целых стран и, соответственно, не способны объяснить в высокой мере дисперсию введенных регрессоров. Поэтому для наблюдения более устойчивых зависимостей вводится ряд контрольных переменных:

INV – объем инвестиций в основной капитал в постоянных ценах по ППС, млн евро (источник: Евростат);

EMPL – численность занятых, тыс. чел. (источник: Евростат);

ESI – индекс экономической сложности (источник: ОЕС [35]);

EXSOC – доля социальных расходов в ВВП, % (источник: Евростат);

HDI – Индекс развития человеческого потенциала, ед. (источник: UNDP [36]).

По всем показателям используются данные за 2016 и 2018 гг. или ближайшие доступные даты для каждой из 29 европейских стран. Формат выборки наблюдений детерминирует использование панельных регрессий: объединенная модель (Pooled model), а также модели с фиксированными (Fixed effect model) и случайными (Random effect model) эффектами. С целью определения наиболее релевантной модели при значимости коэффициентов в разных спецификациях применены следующие тесты: стандартный F-тест, тест Хаусмана и тест Бройша-Пагана [37].

В таблице 2 (по расчетам авторов) представлены наилучшие формы моделей согласно указанному набору тестов, а также скорректированный коэффициент детерминации (R_{adj}^2), размер выборки (N), который варьировался в зависимости от доступных данных, в скобках под коэффициентами при переменных приведено значение t-статистики.

Таблица 2. Модели оценки влияния больших данных на экономические и социальные макропараметры

Зависимая переменная	Уравнение регрессии	Тип модели
ВВП	$GDP = 2.2 * INV + 1924.4 * BD$ (14.00) (2.60) Radj2=0.42; N=51	фиксированные эффекты
Производительность труда	$LP = 23.8 + 23.0 * ECI + 3.3 * BD$ (2.15) (3.27) (4.79) Radj2=0.42; N=47	сквозная
ВВП на душу населения	$GDPpc = 9.6 + 7.4 * ECI + 0.5 * ECOMs + 1.2 * BD$ (1.79) (1.79) (2.14) (3.37) Radj2=0.46; N=46	сквозная
Объем высокотехнологичного экспорта	$EXP = -29640.0 + 0.26 * INV + 2887.3 * BD$ (-3.22) (13.49) (3.81) Radj2=0.81; N=50	сквозная
Доля высокотехнологичного экспорта в общем экспорте	$EXPs = 6.4 + 4.2e-05 * INV - 5.0e-04 * EMPL + 0.4 * BD$ (2.65) (2.66) (-2.09) (2.36) Radj2=0.29; N=52	сквозная
Доля доходов от электронной торговли в валовой выручке предприятий	$ECOMs = 9.0 + 9.5 * \ln(ECI) + 0.6 * BD$ (3.91) (5.09) (2.90) Radj2=0.43; N=46	сквозная
Индекс счастья	$INAP = 4.4 + 0.1 * EXSOC + 0.1 * BD$ (13.16) (4.42) (3.69) Radj2=0.43; N=53	сквозная
Индекс цифровой экономики и общества	$DESI = 2675.1 + 1.8e-02 * GDPpc + 89.4 * BD$ (8.63) (2.94) (3.77) Radj2=0.41; N=52	сквозная
Глобальный индекс сетевого взаимодействия	$GCI = 627.4 * HDI + 0.38 * BD$ (8.38) (2.54) Radj2=0.60; N=47	фиксированные эффекты

С одной стороны, построенные модели не могут претендовать на статус инструмента вычисления точных количественных эффектов от использования больших данных в организациях. Для получения более надежных алгоритмов требуется как существенно большая выборка наблюдений, так и более совершенные математические модели. С другой стороны, статистическая значимость всех построенных регрессий позволяет говорить о явном наличии подобных эффектов. С учетом полученных положительных знаков при регрессоре BD есть все основания говорить о позитивном их характере причем в различных областях экономического и социального развития стран Европы.

4 Пилотный опрос экспертов

Учитывая отсутствие официальных статистических источников, по которым можно было бы оценить социально-экономические эффекты от использования технологий работы с большими данными, для проведения пилотного мониторинга в данной предметной области был использован метод экспертного опроса. Опрос проводился по стандартизированной анкете в заочном формате с использованием электронных средств коммуникаций. Пул экспертов, участвовавших в опросе, состоял из 15 исследователей и специалистов по работе с большими данными. Целью опроса являлась (предварительная) оценка перечня показателей концептуальной схемы мониторинга в области неэкономических (в том числе – социальных) эффектов использования технологий для работы с большими данными, а также экспертное определение значений коэффициентов для расчета показателей экономических эффектов.

Из пула опрошенных экспертов оценку коэффициента расчета косвенных эффектов от работы с большими данными в российской экономике дали 8 человек. Среднее значение по обратным косвенным эффектам использования технологий в области работы с большими данными на российскую экономику составило 0,05; по форвардным косвенным эффектам 1,75; по индуцированным косвенным эффектам – 1,52. При этом по большинству показателей разброс оценок экспертов был сравнительно невелик. Так, оценки коэффициента расчета обратных косвенных эффектов варьировались от 0,3 до 0,8; индуцированных косвенных эффектов – 0,8 до

2,1; и только по показателю коэффициента расчета форвардных косвенных эффектов мнения экспертов разделились (оценки варьировались от 0,3 до 5,0).

В целом полученные в рамках опроса значения хорошо соотносятся с данными по странам Европы, полученными в рамках упоминавшегося ранее проекта Европейской комиссии [17]. Так, в среднем для Великобритании и стран ЕС сумма коэффициентов расчета косвенных эффектов от работы с большими данными в 2020 г. составила 4,53. В то же время разброс значений показателя для стран ЕС довольно велик, например, для Хорватии этот показатель был равен 2,28; а для Германии – 5,58. Таким образом, уровень коэффициента расчета косвенных эффектов от работы с большими данными для российской экономики лишь незначительно уступает среднеевропейскому.

Основываясь на полученных экспертных оценках, можно сделать вывод, что на каждый 1 рубль прямых экономических эффектов (т.е. всей совокупной выручки российских организаций от продажи технологий, решений и услуг работы с большими данными) приходится 3,32 рубля косвенных эффектов.

Для расчета конкретных значений показателей косвенных экономических эффектов от использования технологий работы с большими данными необходимо сначала оценить прямые экономические эффекты (объем совокупной выручки российских организаций от продажи технологий, решений и услуг для работы с большими данными) что будет возможно после проведения опроса организаций-поставщиков данных технологий.

Оценку неэкономических эффектов от использования больших данных дали все 15 экспертов. На основе полученных ответов были определены следующие значения для показателей концептуальной схемы мониторинга BD4DE:

- влияние на безопасность жизнедеятельности в бытовой и профессиональной областях – 3,40;
- влияние на повышение качества предоставляемых услуг – 4,13;
- влияние на экологическую ситуацию в стране – 3,60;
- влияние на занятость населения в российской экономике – 3,20.

Таким образом, с точки зрения экспертов наибольшее воздействие использование больших данных оказывает на повышение качества услуг, в меньшей степени – на безопасность жизнедеятельности в бытовой и профессиональной областях.

Стоит отметить, что для показателей «влияние на экологию» и «влияние на занятость населения» значения меньше 3 баллов означали негативное влияние (уменьшение занятости населения и ухудшение экологической ситуации), 3 балла – отсутствие влияния, более 3 баллов – позитивное влияние. На экологию (в сторону улучшения экологической ситуации) технологии для работы с большими данными по мнению экспертов оказывают незначительное влияние, а на занятость населения в российской экономике влияние – минимальное. При оценке влияния технологий работы с большими данными на занятость населения мнения экспертов разделились: трое респондентов оценили, что работа с большими данными может вызвать незначительное сокращение занятости населения, семь респондентов указали на отсутствие влияния на занятость населения, пятеро указали, что использование больших данных приведет к увеличению занятости населения.

Заключение

Оценка реальных и потенциальных эффектов от использования технологий для работы с большими данными на экономику и общество в России является важнейшей задачей, в результатах которой заинтересованы как органы государственной власти, так и бизнес-структуры. Предложенная концептуальная схема мониторинга BD4DE содержит набор из 8 показателей макроуровня, позволяющих всесторонне оценить как экономические эффекты использования данного типа цифровых технологий, так и их влияние на общество и экологическую ситуацию в стране, а также 10 показателей воздействия использования технологий работы с большими данными на работу организаций. Применение регрессионного анализа (на наборах данных из ряда международных источников) показало безусловное влияние технологий для работы с большими данными на социально-экономические макропараметры.

Апробация концептуальной схемы мониторинга BD4DE была проведена с помощью экспертного опроса, по результатам которого были рассчитаны коэффициенты косвенных эффектов, при этом значения этих показателей оказались сравнимы с показателями по странам Евросоюза и Великобритании. По результатам экспертного опроса были получены значения для показателей, характеризующих неэкономические (в том числе социальные) эффекты от использования данного типа цифровых технологий. По мнению экспертов самое существенное влияние работа с большими данными оказывает на улучшение качества предоставляемых услуг (как органами государственной власти, так и частным сектором), тогда как относительно влияния на рост или сокращение занятости населения мнения экспертов разделились.

В дальнейшем для вычисления значения показателей, характеризующих экономические эффекты, а также для корректировки и актуализации уже полученных значений опрос экспертов следует дополнить результатами опроса организаций-производителей, а также потребителей технологий, решений и услуг в области работы с большими данными.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа одного из авторов статьи (Ю.Е. Хохлов) выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Работа одного из авторов статьи (М.А. Юревич) выполнена в рамках государственного задания Правительства РФ Финансовому университету на 2021 г. по теме «Политико-экономические закономерности функционирования и эволюции экономической системы России».

Работа одного из авторов статьи (С.Б. Шапошник) частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. Cuquet M., Fensel A. The societal impact of big data: A research roadmap for Europe // *Technology in Society*. 2018. Vol. 54. P. 74–86. <https://doi.org/10.1016/j.techsoc.2018.03.005>
2. Choi T. M., Wallace S. W., Wang Y. Big data analytics in operations management // *Production and Operations Management*. 2018. Vol. 27. №. 10. P. 1868–1883. <https://doi.org/10.1111/poms.12838>
3. Wang Y., Kung L. A., Byrd T. A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations // *Technological Forecasting and Social Change*. 2018. Vol. 126. P. 3–13. <https://doi.org/10.1016/j.techfore.2015.12.019>
4. Boyd D., Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon // *Information, communication & society*. 2012. Vol. 15. №. 5. P. 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
5. Jiang M., Fu, K. W. Chinese social media and big data: Big data, big brother, big profit? // *Policy & Internet*. 2018. Vol.10. Iss. 4. P. 372–392. <https://doi.org/10.1002/poi3.187>
6. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
7. Ершова Т.В., Хохлов Ю.Е., Шапошник С.Б. Методология мониторинга развития и использования технологий работы с большими данными // *Информационное общество*, 2021, № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
8. OECD Guide to Measuring the Information Society 2011. URL: <https://www.oecd.org/sti/ieconomy/oecdguidetomeasuringtheinformationsociety2011.htm>

9. Eurostat model for a Community Survey on ICT Usage and e-Commerce in Enterprises. 2008. URL: <https://circabc.europa.eu/sd/a/c2800d69-06e4-4b9b-be18-50a42e4af6d5/Questionnaire%20ENT2008.pdf>
10. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf>
11. OECD. Measuring the Digital Transformation: A Roadmap for the Future. 2019. URL: <https://www.oecd-ilibrary.org/sites/9789264311992-en/index.html?itemId=/content/publication/9789264311992-en> (дата обращения: 01.06.2021)
12. OECD. A Roadmap toward a Common Framework for Measuring the Digital Economy. Report for the G20 Digital Economy Task Force. Saudi Arabia, 2020. URL: <http://www.oecd.org/sti/roadmap-toward-a-common-framework-for-measuring-the-digital-economy.pdf> (дата обращения: 01.06.2021).
13. Bureau of Economic Analysis. Defining and Measuring the Digital Economy. Working Paper. 2018. URL: <https://www.bea.gov/system/files/papers/WP2018-4.pdf> (дата обращения: 31.03.2021).
14. Statistics Canada. Measuring digital economic activities in Canada: Initial estimates. 2019. URL: <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00002-eng.htm> (дата обращения: 31.03.2021).
15. Australian Bureau of Statistics. Measuring digital activities in the Australian economy. URL: <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/ABS+Chief+Economist+-+Full+Paper+of+Measuring+Digital+Activities+in+the+Australian+Economy> (дата обращения: 31.03.2021).
16. Hogan O., Holdgate L., Jayasuriya R. Report for SAS. The Value of Big Data and the Internet of Things to the UK Economy. London, 2016. URL: https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf (дата обращения: 31.03.2021).
17. Cattaneo G, Micheletti G., Glennon M, La Croce C., Mitta C. The European Data Market Monitoring Tool. Key facts & figures, first policy conclusions, data landscape and quantified stories. Final study report. Luxembourg: Publications Office of the European Union, 2020. URL: <https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update> (дата обращения: 31.03.2021).
18. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European Community. Luxembourg: Publications Office of the European Union, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (дата обращения: 31.03.2021).
19. di Bella E., Leporatti L., Maggino F. Big data and social indicators: Actual trends and new perspectives // Social Indicators Research. 2018. Vol. 135. №. 3. P. 869–878.
20. Gupta M., George J. F. Toward the development of a big data analytics capability // Information & Management. 2016. Vol. 53. №. 8. P. 1049–1064.
21. Wamba S. F. et al. Big data analytics and firm performance: Effects of dynamic capabilities // Journal of Business Research. 2017. Vol. 70. P. 356–365.
22. Brynjolfsson E., Hitt L. M., Kim H. H. Strength in numbers: How does data-driven decisionmaking affect firm performance? // SSRN 1819486. 2011. <https://doi.org/10.2139/ssrn.1819486>
23. Brynjolfsson E., McElheran K. Data in action: data-driven decision making in US manufacturing // US Census Bureau Center for Economic Studies Paper No. CES-WP-16-06. Rotman School of Management Working Paper. 2016. №. 2722502.
24. Wu L., Lou B., Hitt L. Data analytics supports decentralized innovation // Management Science. 2019. Vol. 65. №. 10. P. 4863–4877.
25. Tambe P. Big data investment, skills, and firm value // Management Science. 2014. Vol. 60. №. 6. P. 1452–1469.
26. Niebel T., Rasel F., Viete S. BIG data–BIG gains? Understanding the link between big data analytics and innovation // Economics of Innovation and New Technology. 2019. Vol. 28. №. 3. P. 296–316.
27. Симачев Ю. В., Кузык М. Г., Федюнина А. А., Юревич М. А., Зайцев А. А. Факторы роста производительности труда на предприятиях несырьевых секторов российской экономики

- // В кн.: XXI Апрельская международная научная конференция по проблемам развития экономики и общества. М.: Издательский дом НИУ ВШЭ, 2020. С. 1–60.
28. Wu L., Hitt L., Lou B. Data analytics, innovation, and firm productivity // *Management Science*. 2020. Vol. 66. №. 5. P. 2017–2039.
 29. Goodridge P. et al. How does big data affect GDP? Theory and evidence for the UK. Working Papers 25156. London, 2015. URL: <https://ideas.repec.org/p/imp/wpaper/25156.html> (дата обращения: 31.03.2021).
 30. Crépon V., Duguet E., Mairessec J. Research, Innovation and Productivity: An Econometric Analysis at the Firm Level // *Economics of Innovation and new Technology*. 1998. Vol. 7. №. 2. P. 115-158.
 31. Eurostat Database. URL: <https://ec.europa.eu/eurostat/data/database> (дата обращения: 09.04.2021).
 32. SDSN. World Happiness Report 2020. New York: Sustainable Development Solutions Network, 2020. URL: <https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf> (дата обращения: 09.04.2021).
 33. European Commission. Digital Economy and Society Index. URL: <https://digital-agenda-data.eu/datasets/desi/visualizations> (дата обращения: 09.04.2021).
 34. Huawei Technologies Co. GCI Ranking Table. URL: <https://www.huawei.com/minisite/gci/en/country-rankings.html> (дата обращения: 09.04.2021).
 35. ОЕС. Economic Complexity Legacy Rankings (ЕCI). URL: <https://oec.world/en/rankings/country/eci/> (дата обращения: 09.04.2021).
 36. United Nations Development Programme. Human Development Data Center. URL: <http://hdr.undp.org/en/data> (дата обращения: 09.04.2021).
 37. Torres-Reyna O. Getting started in fixed/random effects models using R // *Data & Statistical Services*. Princeton University. 2010. URL: <https://www.princeton.edu/~otorres/Panel101R.pdf> (дата обращения: 09.04.2021).

BIG DATA: SOCIAL AND ECONOMIC EFFECTS

Malakhov, Vadim Aleksandrovich

Candidate of historical sciences

Vavilov Institute for the History of Science and Technology of the Russian Academy of Sciences, Center for the history of science organization and scientific research, senior researcher

Moscow, Russia

yasonbh@mail.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor

Moscow, Russia

yuri.hohlov@iis.ru

Shaposhnik, Sergei Borisovich

Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher

Petrozavodsk, Russian Federation

sergei.shaposhnik@gmail.com

Yurevich, Maksim Andreevich

Financial University under the Government of the Russian Federation, Center for macroeconomic research, researcher

Moscow, Russia

mayurevich@fa.ru

Abstract

A conceptual framework and a set of indicators for monitoring the economic and non-economic effects of big data technologies were developed. The conceptual framework has two main blocks of indicators: the first deals with the impact of big data technologies on the country's economy, the second – on other areas (society, ecology, security). Testing of the conceptual monitoring framework BD4DE was carried out by means of expert survey, according to the results of which indirect effects coefficients were calculated; the values of these indicators were comparable with the indicators of the EU countries and the UK. It was found that the most significant effect of working with big data is to improve the quality of services provided (both by public authorities and the private sector), while experts' opinions were divided regarding the impact on the growth or reduction of employment.

Keywords

big data; big data market; digital economy, big data technologies; monitoring and evaluation of digital technologies; Big Data for Digital Economy; BD4DE; social effect; economic effect; social impact; economic impact

References

1. Cuquet M., Fensel A. The societal impact of big data: A research roadmap for Europe // *Technology in Society*. 2018. Vol. 54. P. 74–86. <https://doi.org/10.1016/j.techsoc.2018.03.005>.
2. Choi T. M., Wallace S. W., Wang Y. Big data analytics in operations management // *Production and Operations Management*. 2018. Vol. 27. №. 10. P. 1868–1883. <https://doi.org/10.1111/poms.12838>.
3. Wang Y., Kung L. A., Byrd T. A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations // *Technological Forecasting and Social Change*. 2018. Vol. 126. P. 3–13. <https://doi.org/10.1016/j.techfore.2015.12.019>.
4. Boyd D., Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon // *Information, communication & society*. 2012. Vol. 15. №. 5. P. 662–679. <https://doi.org/10.1080/1369118X.2012.678878>.
5. Jiang M., Fu, K. W. Chinese social media and big data: Big data, big brother, big profit? // *Policy & Internet*. 2018. Vol.10. Iss. 4. P. 372–392. <https://doi.org/10.1002/poi3.187>.

6. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
7. Ershova T.V., Hohlov Yu.E., Shaposhnik S.B. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // Informatsionnoye obshchestvo, 2021, №4-5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
8. OECD Guide to Measuring the Information Society 2011. URL: <https://www.oecd.org/sti/ieconomy/oecdguidetomeasuringtheinformationsociety2011.htm>
9. Eurostat model for a Community Survey on ICT Usage and e-Commerce in Enterprises. 2008. URL: <https://circabc.europa.eu/sd/a/c2800d69-06e4-4b9b-be18-50a42e4af6d5/Questionnaire%20ENT2008.pdf>
10. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf>
11. OECD. Measuring the Digital Transformation: A Roadmap for the Future. 2019. URL: <https://www.oecd-ilibrary.org/sites/9789264311992-en/index.html?itemId=/content/publication/9789264311992-en> (accessed: 01.06.2021)
12. OECD. A Roadmap toward a Common Framework for Measuring the Digital Economy. Report for the G20 Digital Economy Task Force. Saudi Arabia, 2020. URL: <http://www.oecd.org/sti/roadmap-toward-a-common-framework-for-measuring-the-digital-economy.pdf> (accessed on 31.03.2021).
13. Bureau of Economic Analysis. Defining and Measuring the Digital Economy. Working Paper. 2018. URL: <https://www.bea.gov/system/files/papers/WP2018-4.pdf> (accessed: 31.03.2021).
14. Statistics Canada. Measuring digital economic activities in Canada: Initial estimates. 2019. URL: <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00002-eng.htm> (accessed on 31.03.2021).
15. Australian Bureau of Statistics. Measuring digital activities in the Australian economy. URL: <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/ABS+Chief+Economist+-+Full+Paper+of+Measuring+Digital+Activities+in+the+Australian+Economy> (accessed on 31.03.2021).
16. Hogan O., Holdgate L., Jayasuriya R. Report for SAS. The Value of Big Data and the Internet of Things to the UK Economy. London, 2016. URL: https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf (accessed on 31.03.2021).
17. Cattaneo G, Micheletti G., Glennon M, La Croce C., Mitta C. The European Data Market Monitoring Tool. Key facts & figures, first policy conclusions, data landscape and quantified stories. Final study report. Luxembourg: Publications Office of the European Union, 2020. URL: <https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update> (accessed on 31.03.2021).
18. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European Community. Luxembourg: Publications Office of the European Union, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (accessed on 31.03.2021).
19. di Bella E., Leporatti L., Maggino F. Big data and social indicators: Actual trends and new perspectives // Social Indicators Research. 2018. Vol. 135. №. 3. P. 869–878.
20. Gupta M., George J. F. Toward the development of a big data analytics capability // Information & Management. 2016. Vol. 53. №. 8. P. 1049–1064.
21. Wamba S. F. et al. Big data analytics and firm performance: Effects of dynamic capabilities // Journal of Business Research. 2017. Vol. 70. P. 356–365.
22. Brynjolfsson E., Hitt L. M., Kim H. H. Strength in numbers: How does data-driven decisionmaking affect firm performance? // SSRN 1819486. 2011. <https://doi.org/10.2139/ssrn.1819486>
23. Brynjolfsson E., McElheran K. Data in action: data-driven decision making in US manufacturing // US Census Bureau Center for Economic Studies Paper No. CES-WP-16-06, Rotman School of Management Working Paper. 2016. №. 2722502.
24. Wu L., Lou B., Hitt L. Data analytics supports decentralized innovation // Management Science. 2019. Vol. 65. №. 10. P. 4863–4877.
25. Tambe P. Big data investment, skills, and firm value // Management Science. 2014. Vol. 60. №. 6. P. 1452–1469.

26. Niebel T., Rasel F., Viete S. BIG data–BIG gains? Understanding the link between big data analytics and innovation // *Economics of Innovation and New Technology*. 2019. Vol. 28. №. 3. P. 296–316.
27. Simachev YU. V., Kuzyk M. G., Fedyunina A. A., Yurevich M. A., Zajtsev A. A. Faktory rosta proizvoditel'nosti truda na predpriyatiyakh nesyr'evykh sektorov rossijskoj ehkonomiki // V kn.: XXI Aprel'skaya mezhdunarodnaya nauchnaya konferentsiya po problemam razvitiya ehkonomiki i obshhestva. M. : Izdatel'skij dom NIU VSHE, 2020. P. 1–60.
28. Wu L., Hitt L., Lou B. Data analytics, innovation, and firm productivity // *Management Science*. 2020. Vol. 66. №. 5. P. 2017–2039.
29. Goodridge P. et al. How does big data affect GDP? Theory and evidence for the UK. Working Papers 25156. London, 2015. URL: <https://ideas.repec.org/p/imp/wpaper/25156.html> (accessed on 31.03.2021).
30. Crépon B., Duguet E., Mairessec J. Research, Innovation and Productivity: An Econometric Analysis at the Firm Level // *Economics of Innovation and new Technology*. 1998. Vol. 7. №. 2. P. 115–158.
31. Eurostat Database. URL: <https://ec.europa.eu/eurostat/data/database> (accessed: 09.04.2021).
32. SDSN. World Happiness Report 2020. New York: Sustainable Development Solutions Network, 2020. URL: <https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf> (accessed on 09.04.2021).
33. European Commission. Digital Economy and Society Index. URL: <https://digital-agenda-data.eu/datasets/desi/visualizations> (accessed on 09.04.2021).
34. Huawei Technologies Co. GCI Ranking Table. URL: <https://www.huawei.com/minisite/gci/en/country-rankings.html> (accessed: 09.04.2021).
35. OEC. Economic Complexity Legacy Rankings (ECI). URL: <https://oec.world/en/rankings/country/eci/> (accessed on 09.04.2021).
36. United Nations Development Programme. Human Development Data Center. URL: <http://hdr.undp.org/en/data> (accessed on 09.04.2021)
37. Torres-Reyna O. Getting started in fixed/random effects models using R // *Data & Statistical Services*. Princeton University. 2010. URL: <https://www.princeton.edu/~otorres/Panel101R.pdf> (accessed on 09.04.2021).

Государственное управление в информационном обществе

МОНИТОРИНГ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ В СИСТЕМЕ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ РОССИИ

Катин Александр Владимирович

*Институт развития информационного общества, руководитель дирекции отраслевых программ РЭУ имени Г. В. Плеханова, старший преподаватель базовой кафедры цифровой экономики ИРИО Москва, Российская Федерация
alexander.katin@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Работа посвящена исследованию вопросов мониторинга и оценки уровня использования технологий работы с большими данными в системе государственного управления в России. Для разработки методологии и концептуальной схемы мониторинга, а также выбора соответствующих показателей был проведен анализ релевантных научных работ, проиндексированных платформой Web of Science за последние 5 лет, а также ведущих международных российских и международных организаций, занимающихся исследованием вопросов мониторинга использования цифровых технологий в системе государственного управления. Результаты проведенного анализа показали, что на сегодня отсутствуют единые, признанные всеми подходы к оценке и мониторингу использования технологий работы с большими данными в органах государственной власти. В то же время тема является достаточно актуальной, что послужило импульсом к проведению настоящей работы.

Ключевые слова

большие данные, технологии работы с большими данными, государственное управление, органы власти, уровень зрелости

Введение

За последние годы государство сделало большие инвестиции в развитие информационных систем. Количество таких систем, массив государственных данных, интенсивность обмена ими постоянно растут. По оценкам Минцифры России, по состоянию на конец мая 2019 года в государственных информационных системах хранится более 80 Пб информации. При этом в течение следующих пяти лет ожидается удвоение этого объема [1].

Использование технологий хранения и анализа больших данных в системе государственного и муниципального управления становится важным фактором эффективной реализации функций и полномочий системы государственного управления, а также обеспечения высокого уровня качества и доступности услуг, оказываемых гражданам и организациям. Это подтверждается аналитическими докладами, выпущенными авторитетными международными организациями или международными консалтинговыми компаниями (см. Всемирный Банк [2], Ernst & Young [3], SAS Institute [4]). Организация экономического сотрудничества и развития (далее ОЭСР) в своих докладах [5, 6] также уделяет много внимания цифровой трансформации системы

© Катин А. В., Хохлов Ю. Е., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

DOI: 10.52605/16059921_2021_04_150

государственного управления, фокусируясь на использовании всех преимуществ от использования данных.

Мониторинг уровня использования технологий работы с большими данными в государственном и муниципальном управлении является важным направлением деятельности, поскольку дает возможность оценить вклад уже принятых мер и вложенных ресурсов, подтвердить или опровергнуть адекватность поставленных целей задач и показателей эффективности, и, в случае необходимости, их скорректировать.

Целью данной статьи является разработка компонентов концептуальной схемы мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации (Big Data for Digital Economy, BD4DE), относящихся к методологии проведения мониторинга и оценки уровня использования технологий работы с большими данными в системе государственной управления.

1 Предметная область мониторинга и обзор литературы

1.1 Определение предметной области мониторинга

Предметной областью мониторинга определена деятельность органов государственной власти и органов местного самоуправления по использованию технологий работы с большими данными для всех уровней власти: федерального, регионального и местного. Под деятельностью будем понимать оказание государственных (муниципальных) услуг, исполнение государственных (муниципальных) функций, а также реализацию других полномочий, возложенных на орган государственной власти и органов местного самоуправления в соответствии с законодательством и подзаконными актами.

На федеральном уровне мониторинг должен проводиться в разрезах федеральных органов исполнительной власти и государственных внебюджетных фондов. Для региональных органов исполнительной власти оценка должна осуществляться в разрезах субъектов Российской Федерации. Мониторинг и оценка уровня использования технологий работы с большими данными органами органов местного самоуправления должны проводиться в разрезах поселений, муниципальных районов, муниципальных округов, городских округов, городских округов с внутригородским делением и внутригородских районов.

Следует отметить, что на использование технологий работы с большими данными в системе государственного управления влияет множество факторов – таких как государственная политика и регулирование, развитость системы управления и финансирования, лидерство и человеческий капитал, готовность цифровой инфраструктуры и уровень обеспечения безопасности и доверия. Также важным элементом мониторинга является отслеживание получаемых эффектов (социальных и экономических) от использования в системе государственного управления технологий работы с большими данными. Подходы к мониторингу и оценке перечисленных факторов и эффектов определены в отдельных статьях настоящего журнала, в то время как в настоящей работе основной упор сделан на мониторинг и оценку уровня использования технологий работы с большими данными органами государственной власти и органами местного самоуправления.

1.2 Обзор литературы

1.2.1 Обзор научных публикаций

Для формирования концептуальной схемы мониторинга и оценки уровня использования технологий для работы с большими данными в системе государственного управления в России было проведено исследование наиболее релевантных теме исследования научных публикаций, индексируемых платформой Web of Science.

На первом этапе был проведен отбор публикаций за временной период с 2016 по 2020 годы по поисковому образцу (представлен ниже), сформированному для выявления работ, относящихся к тематике больших данных, процесс разработки которого подробно описан в работе [7].

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Поиск проводился по названию, аннотации, автору и ключевым словам (применено поле «Тема»). В результате было отобрано 108 073 статьи.

На следующем этапе отбора публикаций был разработан специализированный поисковый образ, позволяющий отобрать публикации, относящиеся к системе государственного управления:

“public administration” OR “public authority” OR “public service*” OR “public sector” OR “public function*” OR “digital government” OR “digital government maturity” OR “e-government” OR “electronic government” OR “e-government maturity” OR “digital public service” OR municipal OR “government digitalization” OR “local government” OR “public governance” OR “data-driven public sector” OR “user-driven public sector” OR “citizen-driven public sector” OR “data-driven government” OR e-participation OR e-inclusion OR “citizen-centricity” OR “government effectiveness” OR “government efficiency” OR “open government” OR “open government data” OR “legislative branch” OR “executive branch” OR “judicial branch”

Подбор ключевых слов на специализированного поискового образа осуществлялся методом экспертных оценок с последующей верификацией, что позволило максимально охватить исследуемую отрасль и исключить из полученной выборки значительное количество нерелевантных теме публикации исследований.

Общий поисковый запрос был объединен со специализированным посредством логического оператора “AND”, в результате чего на платформе Web of Science было найдено 1294 публикации.

Во всех отобранных публикациях были проанализированы аннотации на предмет наиболее близкого соответствия теме исследования, в результате чего было выделено 30 работ [8-37], для которых было принято решение изучить полный текст.

Часть отобранных работ [8-13] посвящена исследованию вопросов готовности органов власти к использованию технологий работы с большими данными. Например, в работе [8] приводится обзор моделей зрелости работы с большими данными для государственного сектора, а в работе [13] рассмотрены вопросы зрелости процессов управления данными в системе государственного управления.

Еще одним блоком вопросов, которым посвящены отобранные публикации, являются статьи, исследующие процессы внедрения и использования технологий работы с большими данными в органах государственной власти и органах местного самоуправления [14-26]. В частности, в работе [15] рассмотрены вопросы применения инструментов прогнозной аналитики в процессах принятия решений в системе государственного управления. В статье [22] проводится широкий обзор использования технологий работы с большими данными в системе государственного управления.

Важной темой, которой посвящено значительное количество отобранных публикаций [27-37], являются препятствия для внедрения и использования технологий работы с большими данными, стоящие перед организациями системы государственного управления. Например, в работе [32] приведен общий обзор барьеров для использования технологий работы с большими данными в системе государственного управления, а статья [27] в основном посвящена рискам, связанным с информационной безопасностью применения указанных технологий в органах государственной власти и органов местного самоуправления.

Анализ отобранных научных публикаций позволяет установить, что в настоящее время отсутствуют всеми признанные подходы к мониторингу и оценке уровня использования технологий работы с большими данными в системе государственного управления. В то же время все работы, приведенные выше, можно разделить на три блока: исследования готовности органов власти к использованию технологий работы с большими данными, барьеры для их полноценного использования, а также обзор и описания практик применения отдельных решений и инструментов, основанных на больших данных, в системе государственного управления. Выделенные блоки могут быть использованы в качестве основы для формирования концептуальной схемы и выделения отдельных показателей мониторинга и оценки уровня использования технологий работы с большими данными в системе государственного управления в России.

1.2.2 Обзор аналитических материалов

Как уже говорилось во введении, ряд ведущих аналитических компаний и организаций занимается вопросами мониторинга и оценки уровня использования системой государственного управления цифровых технологий в целом и технологий работы с большими данными в частности.

В 2017 году Всемирным банком в сотрудничестве с Институтом развития информационного общества была разработана методика оценки цифровой экономики, предназначенная для различных стран мира [38], которая включает концептуальную схему, содержащую следующие элементы:

- факторы (основы) развития цифровой экономики;
- уровень использования цифровых технологий для трансформации ключевых сфер деятельности (государственного сектора, бизнеса), а также домохозяйствами и населением;
- воздействие цифровых технологий на социально-экономическое развитие (экономический рост, рабочие места, качество услуг).

Организацией экономического сотрудничества и развития регулярно публикуются доклады по Индексу цифрового правительства [39], в основе которого лежит концептуальная схема, содержащая оценку следующих принципов «полностью цифрового правительства»:

- цифровое правительство по умолчанию;
- правительство как платформа;
- управление государственным сектором, основанное на данных;
- открытое правительство по умолчанию;
- правительство, управляемое пользователями;
- проактивное правительство.

Некоторые элементы приведенных работ могут быть использованы при формировании концептуальной схемы мониторинга и оценки уровня использования технологий для работы с большими данными в системе государственного управления, поскольку данным в них уделяется особое внимание.

2 Концептуальная схема и показатели мониторинга

Исходя из определенной предметной области мониторинга, а также из проведенного анализа релевантных источников, концептуальная схема мониторинга и оценки использования технологий работы с большими данными (представлена на рисунке 1) включает в себя три области: уровень зрелости работы с большими данными в системе государственного управления и в органах местного самоуправления, использование технологий работы с большими данными органами государственной власти и органами местного самоуправления, а также барьеры для использования этих технологий.



Рисунок 1. Концептуальная схема мониторинга и оценки уровня использования технологий работы с большими данными в системе государственного управления в России

2.1 Уровень зрелости работы с большими данными в системе государственного управления

Оценка уровня зрелости работы с большими данными позволяет охарактеризовать качество процессов и действий внутри органа государственной власти или органа местного самоуправления, связанная с хранением и обработкой больших данных. Оценка осуществляется с использованием модели зрелости, подробно описанной в работе [40]. Данный инструмент дает возможность оценить текущее состояние работы с большими данными в органе государственной власти или органе местного самоуправления, выявить слабые и сильные стороны и выстроить план действий по повышению эффективности функционирования организации. Модель зрелости включает 5 уровней зрелости. Наиболее низкий («начальный») уровень зрелости позволяет сделать вывод, что на текущий момент организация практически не работает с большими данными. «Высокий» уровень могут получить органы государственной власти и органы местного самоуправления, деятельность которых в наибольшей степени опирается на результаты работы с большими данными.

Модель зрелости включает следующие размерности: стратегия и регулирование; кадры и лидерство; данные; инструменты и аналитика; инфраструктура и безопасность; организация работы с большими данными; воздействие работы с большими данными.

- (УЗГУ-01) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень зрелости с точки зрения работы с большими данными

Показатель позволяет дать интегральную оценку уровня зрелости работы с большими данными в соответствии с примененной моделью для всей системы государственного управления в России.

- (УЗГУ-01-01) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень зрелости с точки зрения наличия стратегии и регулирования работы с большими данными

Показатель дает возможность отслеживать степень развитости в органах государственной власти и органах местного самоуправления процессов стратегического планирования, а также нормативного правового и технического регулирования использования технологий работы с большими данными в своей деятельности.

- (УЗГУ-01-02) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень кадрового обеспечения работы с большими данными

Наличие необходимых знаний, навыков и компетенций у кадров является ключевым фактором успешной работы с большими данными в организации. Показатель позволяет оценить долю органов государственной власти и органов местного самоуправления, имеющих достаточный уровень развитости кадрового обеспечения для работы с большими данными.

- (УЗГУ-01-03) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень обеспеченности качественными большими данными

Полноценная работа с большими данными невозможна без достижения достаточного уровня качества данных. Показатель позволяет оценить, насколько российская система государственного управления обеспечена качественными большими данными.

- (УЗГУ-01-04) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень использования аналитики и инструментов для работы с большими данными

Показатель позволяет проводить оценку уровня использования аналитики и инструментов для работы с большими данными в различные виды деловых административных процессов органов государственной власти и органов местного самоуправления.

- (УЗГУ-01-05) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень зрелости с точки зрения цифровой инфраструктуры и безопасности для работы с большими данными

Использование показателя даст возможность отслеживания развитости цифровой инфраструктуры, необходимой для работы с большими данными в органах государственной власти и органах местного самоуправления, а также возможность оценки достигаемого уровня защищенности при работе с большими данными, включая мониторинг используемых специализированных средств обеспечения информационной безопасности.

- (УЗГУ-01-06) Доля органов государственной власти и органов местного самоуправления, имеющих продвинутый и выше уровень зрелости организационных механизмов работы с большими данными

Деятельность органов государственной власти и органов местного самоуправления зачастую достаточно сильно зарегулирована и должна соответствовать имеющимся регламентам, что может мешать полноценному использованию технологий работы с большими данными. Показатель позволяет оценить долю организаций российской системы государственного управления, имеющих продвинутый уровень организационных механизмов работы с большими данными.

2.2 Использование технологий работы с большими данными

Мониторинг уровня использования технологий хранения и анализа больших данных в государственном и муниципальном управлении позволяет осуществлять оценку проникновения указанных технологий в повседневное функционирование системы государственного управления России.

Сегодня в Российской Федерации отсутствуют показатели, которые характеризовали бы уровень использования технологий работы с большими данными в государственном и муниципальном управлении. Исходя из проведенного обзора релевантных научных публикаций и аналитических материалов, а также реализованных ранее проектов, были определены показатели, которые дают возможность оценить интенсивность использования технологий сбора, хранения и анализа больших данных в органах государственной власти и органах местного самоуправления Российской Федерации. Преимуществом выбранных показателей также является возможность для отдельных из них осуществлять сопоставление с данными, собираемыми в ОЭСР для предприятий [41].

- (ИБДГУ-01) Доля органов государственной власти и органов местного самоуправления, использующих технологии работы с большими данными

Показатель позволит оценить степень проникновения технологий работы с большими данными в деятельность органов государственной власти и органов местного самоуправления. Важным элементом данного показателя является возможность оценки с точки зрения различных этапов жизненного цикла данных, а именно их генерации, хранения, анализа и визуализации.

- (ИБДГУ-02) Доля органов государственной власти и органов местного самоуправления, осуществляющих работу с большими данными собственными силами или силами внешних компаний

Важным направлением мониторинга является отслеживание возможностей органов государственной власти и органов местного самоуправления с точки зрения самостоятельной работы с большими данными. Для этого введен показатель, который позволяет оценить долю органов государственной власти и органов местного самоуправления, осуществляющих анализ больших данных собственными силами, а также силами внешних компаний.

- (ИБДГУ-03) Доля органов государственной власти и органов местного самоуправления, предоставляющих (продающих), а также получающих (покупающих) доступ к большим данным

Для полноценного понимания процессов работы с большими данными в органах государственной власти и органах местного самоуправления, необходимо установить, получают ли государственные организации доступ к большим данным частных компаний и предоставляют ли они доступ к государственным данным частным организациям. В силу законодательных ограничений предоставление доступа частным организациям к государственным данным достаточно проблематично, однако возможно, например, в рамках государственно-частного партнерства.

- (ИБДГУ-04) Доля затрат органов государственной власти и органов местного самоуправления на работу с большими данными в общих затратах на цифровые технологии

Крайне важным направлением мониторинга является оценка уровня затрат органов государственной власти и органов местного самоуправления на технологии работы с большими данными. Для это введен показатель, который даст возможность оценить долю расходов на эти технологии в общих затратах на ИКТ органов государственной власти и органов местного самоуправления.

- (ИБДГУ-05) Доля органов государственной власти и органов местного самоуправления, применяющих аналитику больших данных (в разрезах по целям использования)

Показатель позволит отслеживать возможности российских органов государственной власти и органов местного самоуправления с точки зрения применения аналитики данных для различных целей, среди которых:

- исполнение государственных (муниципальных) функций;
- оказание государственных (муниципальных услуг);
- управление организацией;
- обеспечение информационной безопасности;
- управление кадрами.

2.3 Барьеры для использования технологий работы с большими данными

Мониторинг основных барьеров для внедрения и использования технологий работы с большими данными в государственном и муниципальном управлении позволяет установить основные причины, в связи с которыми организации отказываются от работы с технологиями. Понимание главных препятствий позволит выявить слабые места органов государственной власти и органов местного самоуправления и повысить эффективность процессов цифровой трансформации за счет выстраивания правильных приоритетов в этой деятельности.

- (ББДГУ-01) Доля органов государственной власти и органов местного самоуправления, считающих, что затраты на технологии работы с большими данными слишком высоки

Одной из основных причин неиспользования технологий работы с большими данными является ограниченность финансовых ресурсов; следовательно, необходимо оценить долю органов государственной власти и органов местного самоуправления, считающих, что затраты на технологии работы с большими данными слишком высоки по сравнению с выгодами, которые потенциально могут быть получены.

- (ББДГУ-02) Доля органов государственной власти и органов местного самоуправления, считающих, что у них недостаточно человеческих ресурсов, знаний и навыков

Еще одним барьером является нехватка квалифицированных кадров, особенно характерная для органов государственной власти и органов местного самоуправления в силу неконкурентного уровня заработной платы по сравнению с частным сектором. В рамках мониторинга готовности к использованию технологий работы с большими данными необходимо оценить долю органов государственной власти и органов местного самоуправления, считающих, что у них недостаточно человеческих ресурсов, знаний и навыков.

- (ББДГУ-03) Доля органов государственной власти и органов местного самоуправления, считающих, что их цифровая инфраструктура недостаточно развита для использования технологий работы с большими данными

Критическим препятствием для полноценного внедрения и использования технологий работы с большими данными является отсутствие в органах государственной власти и органах местного самоуправления адекватной цифровой инфраструктуры. В процессе мониторинга необходимо оценить долю организаций, считающих, что их цифровая инфраструктура недостаточно развита для использования технологий работы с большими данными, которые требуют высокой скорости и надежности интернет-соединения, больших вычислительных мощностей и т.д.

- (ББДГУ-04) Доля органов государственной власти и органов местного самоуправления, считающих, что технологии работы с большими данными не являются приоритетом для их деятельности

Применение технологий хранения и анализа больших данных является относительно новым для органов государственной власти и органов местного самоуправления в Российской Федерации. Показатель позволяет оценить долю организаций системы государственного управления в России, не использующих технологии работы с большими данными, так как это не является для них приоритетом.

- (ББДГУ-05) Доля органов государственной власти и органов местного самоуправления, считающих, что источников больших данных для решения их текущих задач, недостаточно

Полноценное функционирование технологий работы с большими данными и достижение положительных эффектов от их использования невозможно без достаточного количества различных источников данных. Показатель позволяет оценить долю органов государственной

власти и органов местного самоуправления, считающих, что таких источников на текущий момент не хватает.

- (ББДГУ-06) Доля органов государственной власти и органов местного самоуправления, считающих, что качество имеющихся больших массивов данных (как в организации, так и за ее пределами) не позволяет решать текущие задачи

Наряду с наличием разнообразных источников данных, а также их необходимого объема, важным условием для использования технологий работы с большими данными является качество данных. Показатель позволяет оценить долю организаций системы государственного управления, не использующих технологии работы с большими данными по причине недостаточного уровня качества данных.

- (ББДГУ-07) Доля органов государственной власти и органов местного самоуправления, полагающих, что технологии работы с большими данными бесполезны

Помимо наличия таких значимых факторов, как развитая цифровая инфраструктура и квалифицированные кадры, важным аспектом процесса внедрения технологий работы с большими данными в деятельность органов государственной власти и органов местного самоуправления является наличие у лиц, принимающих решения, информированности о преимуществах, которые возможно получить от использования указанных технологий. Для правильного выстраивания информационной политики представляется необходимым установить долю организаций системы государственного управления в России, считающих технологии работы с большими данными бесполезными.

- (ББДГУ-08) Доля органов государственной власти и органов местного самоуправления, отмечающих трудности соблюдения законодательства о персональных данных при работе с большими данными

Органы государственной власти в силу специфики своей деятельности оперируют большими объемами персональных данных, что накладывает значительные законодательные ограничения на процессы их хранения и обработки. Показатель позволяет выяснить долю органов государственной власти и органов местного самоуправления, опасаящихся трудностей в соблюдении законодательства о персональных данных при применении технологий работы с большими данными.

- (ББДГУ-09) Доля органов государственной власти и органов местного самоуправления, считающих использование аналитики больших данных неэтичным

От эффективности деятельности органов государственной власти и органов местного самоуправления напрямую зависит качество жизни граждан и функционирование организаций частного сектора. Следовательно, у всех заинтересованных сторон могут возникать сомнения по поводу этичности принятия тех или иных решений на основе анализа больших данных, например, в части осуществления социальной поддержки населения либо в процессе осуществления контрольно-надзорной деятельности. Показатель позволяет оценить долю органов государственной власти и органов местного самоуправления, считающих использования аналитики больших данных неэтичным.

3 Методология мониторинга

Для части показателей, позволяющих оценить уровень использования технологий работы с большими данными в системе государственного управления, данные будут взяты из результатов федерального статистического наблюдения, проведенного Росстатом по новой форме №3-информ [42], которая содержит соответствующие вопросы. Данные будут доступны осенью 2021 года и позволят рассчитать значения за 2020 год для показателей ИБДГУ-01, ИБДГУ-02 и ИБДГУ-05.

Для расчета значений остальных показателей данные должны будут собираться в рамках выборочных представительных опросов органов государственной власти и органов местного самоуправления. Для отбора органов государственной власти должна использоваться случайная стратифицированная выборка. Подробное описание методологии сбора данных и расчета значений показателей для отдельной сферы деятельности приведено в статье [7].

Опрос может проводиться в зависимости от бюджета методом личных интервью по специализированной анкете или с использованием электронных средств коммуникаций путем

заполнения респондентами онлайн-формы. Предпочтительным является метод личного интервью.

В качестве основного респондента выбирается лицо, принимающее решения и несущее основную ответственность за вопросы, связанные с развитием и использованием ИКТ в органе государственной власти или органе местного самоуправления.

Отчетным периодом является предыдущий календарный год, где это указано (2020 при пилотном и первом опросах). В случаях, когда не указан отчетный период, респонденты должны относить ответы к текущему положению дел (на момент обследования).

Для проведения опроса органов государственной власти и органов местного самоуправления была разработана специализированная анкета, содержащая модельные вопросы по каждому показателю, а также необходимые инструкции по заполнению, включающие пояснения и определения.

Для расчета значений показателей были разработаны соответствующие паспорта, содержащие алгоритмы расчета значений, разрезы и весовые коэффициенты.

4 Результаты пилотной реализации

На первой стадии пилотной реализации системы мониторинга данной предметной области, связанной с использованием технологий работы с большими данными в системе государственного и муниципального управления, а именно стадии экспертной апробации системы показателей, реализованной специалистами по мониторингу развития цифровой экономики и статистики информационного общества, было принято решение уточнить состав используемых показателей для учета дополнительных факторов, влияющих на применение технологий работы с большими данными.

На указанной стадии модель зрелости была адаптирована для применения в органах государственной власти и органах местного самоуправления, прежде всего были уточнены аспекты целеполагания работы с большими данными и возникающие при этом эффекты. Остальные компоненты данной предметной области мониторинга, связанные с уровнем использования технологий работы с большими данными, а также с барьерами для использования этих технологий, было решено оставить практически без изменения.

На второй стадии пилотной реализации системы мониторинга проводилась экспертная оценка доработанной анкеты экспертом в данной предметной области (специалистом по применению цифровых технологий в системе государственного и муниципального управления) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций).

Экспертом по применению цифровых технологий в государственном и муниципальном управлении были сделаны предложения по уточнению формулировок для учета специфики данной предметной области.

На основании официальных статистических данных за 2020 г. [43] были рассчитаны значения для показателей ИБДГУ-01 и, частично, для ИБДГУ-05 (рисунок 2).



Рисунок 2. Доля организаций государственного и муниципального управления, выполнявших анализ больших данных в 2020 году, %.

Доля организаций государственного и муниципального управления, использовавших технологии сбора, обработки и анализа больших данных в 2020 году составила 17.4%. Выполняют анализ больших данных существенно меньше организаций государственного и муниципального управления - 5.5 %, причем 3% осуществляют данную деятельность только собственными силами и всего 0.8% привлекают этого сторонние организации

По итогам проведения пилотной реализации была доработана анкета для обследования органов государственной власти и органов местного самоуправления по вопросам использования технологий работы с большими данными, которая в дальнейшем будет использоваться при проведении опроса органов государственной власти и органов местного самоуправления.

Заключение

В данной работе описана методология мониторинга и оценки уровня использования технологий работы с большими данными в системе государственного управления в России. Концептуальная схема, предлагаемая для использования в рамках методологии, основывается на изучении релевантных научных публикаций, вышедших за последние 5 лет, а также на работах ведущих аналитических и исследовательских международных организаций. Она включает в себя три области: уровень зрелости работы с большими данными в органах государственной власти и органах местного самоуправления, использования указанных технологий в системе государственного управления России, а также оценку основных барьеров на пути их внедрения и эффективного использования.

Применимость предлагаемой концептуальной схемы и вошедших в нее показателей мониторинга доказана в рамках пилотной реализации с учетом рекомендованных доработок. Следующим этапом апробации предложенного подхода должно стать масштабное полевое исследование, включающее в себя опрос органов государственной власти и органов местного самоуправления с использованием разработанной анкеты. В настоящее время данные для расчета значений показателей недоступны.

Дальнейшее развитие настоящего исследования может идти в направлении уточнения концептуальной схемы мониторинга, а также пересмотра перечня показателей, который может потребоваться после сбора всех необходимых данных для проведения оценки, а также получения обратной связи от опрашиваемых организаций и профессионального сообщества.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. Государственные данные — это грязная нефть. Глава Минцифры Носков представил Медведеву план работы с Big Data // Интернет-издание Tadviser. 2019. URL: [https://www.tadviser.ru/index.php/Статья:Национальная_система_управления_данными_\(НСУД\)_](https://www.tadviser.ru/index.php/Статья:Национальная_система_управления_данными_(НСУД)_) (дата обращения: 01.06.2021)
2. World Bank (2017). Big Data in Action for Government: Big Data Innovation in Public Services, Policy, and Engagement. World Bank, Washington, DC. URL: <https://openknowledge.worldbank.org/handle/10986/26391> (дата обращения: 01.06.2021)
3. Ernst and Young. How does digital government become better government? URL: https://www.ey.com/en_gl/government-public-sector/how-does-digital-government-become-better-government (дата обращения: 01.06.2021)

4. SAS Institute. Big data in government: How data and analytics power public programs. URL: https://www.sas.com/en_us/insights/articles/big-data/big-data-government.html (дата обращения: 01.06.2021)
5. OECD (2019), The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>.
6. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government. <https://dx.doi.org/10.1787/f64fed2a-en>
7. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
8. Haddad A., Ameen A., Isaac O., Alrajawy I., Al-Shbami A., Midhun Chakkaravarthy D. (2020) The Impact of Technology Readiness on the Big Data Adoption Among UAE Organisations. In: Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing, vol 1016. Springer, Singapore. https://doi.org/10.1007/978-981-13-9364-8_19
9. Aras Okuyucu, Nilay Yavuz. Big data maturity models for the public sector: a review of state and organizational level models. Transforming Government: People, Process and Policy © Emerald Publishing Limited 1750-6166 <https://doi.org/10.1108/TG-09-2019-0085>
10. Hee Yeong Kim, June-Suh Cho. Data Governance Framework for Big Data Implementation with a Case of Korea. IEEE Computer Society. 978-1-5386-1996-4/17 \$31.00 © 2017 IEEE. <https://doi.org/10.1109/BigDataCongress.2017.56>
11. Ku, M. and Gil-Garcia, J. R. 2018. Ready for Data Analytics? Data Collection and Creation in Local Governments, In Proceedings of the 19th Annual International Conference on Digital Government Research, May 30 – June 1, Delft, the Netherlands <https://doi.org/10.1145/3209281.3209381>
12. Feldiansyah Bakri Nasution, Dr. Nor Erne Nazira Bazin. E-Government Maturity Model to Support System Dynamics in Public Policymaking. Proceeding of EECISI 2018, Malang - Indonesia, 16-18 Oct 2018 (p 465-471). <https://doi.org/10.1109/EECSI.2018.8752921>
13. Manoj A. Thomasa, Joseph Cipollab, Bob Lambertc, Lemuria Carterd. Data management maturity assessment of public sector agencies. Government Information Quarterly, <https://doi.org/10.1016/j.giq.2019.101401>
14. Katarina Tomičić-Pupek, Igor Pihir, Martina Tomičić Furjan. Smart city initiatives in the context of digital transformation – scope, services and technologies. Journal of Contemporary Management Issues . <https://doi.org/10.30924/mjcmi.24.1.3>
15. Tobias Brandt , Sebastian Wagner , Dirk Neumann , Prescriptive Analyticsin Public-Sector Decision-Making: A Framework and Insights from Charging Infrastructure Planning, European Journal of Operational Research (2020), <https://doi.org/10.1016/j.ejor.2020.09.034>
16. Ji Yeon Cho, Bong Gyou Lee. Creating value using public big data: comparison of driving factors from the provider’s perspective. Information Technology & People © Emerald Publishing Limited 0959-3845. <https://doi.org/10.1108/ITP-04-2019-0169>
17. Aakash Puri. Application and Uses of Big Data Predictive Analysis in Public Sectors: A Systematic Review. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (p 539 - 543). <https://doi.org/10.1109/CTEMS.2018.8769196>
18. Sounman Hong , Sun Hyoung Kim, Youngrok Kim and Jeongin Park. Big Data and government: Evidence of the role of Big Data for smart cities. Big Data & Society January–June 2019: 1–11. <https://doi.org/10.1177/2053951719842543>
19. Esmat A. Wahdain, Ahmad Suhaimi Baharudin, and Mohammad Nazir Ahmad. Big Data Analytics in the Malaysian Public Sector: The Determinants of Value Creation. Springer Nature Switzerland AG 2019 F. Saeed et al. (Eds.): IRICT 2018, AISC 843, pp. 139–150, 2019.DOI: https://doi.org/10.1007/978-3-319-99007-1_14
20. Kevin C. Desouza and Benoy Jacob. Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration & Society 1-22. <https://doi.org/10.1177/0095399714555751>
21. Ismagilova E, Hughes L, Dwivedi YK et al (2019) Smart cities: Advances in research – An information systems perspective., International Journal of Information Management. 47: 88-100. <https://doi.org/10.1016/j.ijinfomgt.2019.01.004>

22. Mariusz Maciejewski. To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences* 0(0) 1–16 (2016). <https://doi.org/10.1177/0020852316640058>
23. Mohammad I. Merhi, Klajdi Bregu. Effective and efficient usage of big data analytics in public sector. *Transforming Government: People, Process and Policy* © Emerald Publishing Limited 1750-6166. <https://doi.org/10.1108/TG-08-2019-0083>
24. Liu Naili and Ma Lei. Research and Application of Big Data in E-government. *IOP Conf. Series: Materials Science and Engineering* 563 (2019) 052086. IOP Publishing. <https://doi.org/10.1088/1757-899X/563/5/052086>
25. Kamran Soomro, Nasir Bhutta y, Zaheer Khanz, Muhammad Atif Tahir. Smart City Big Data Analytics. An Advanced Review. *Wires data mining and knowledge discovery*. Volume 9, Issue 5 September/October 2019. <https://doi.org/10.1002/widm.1319>
26. Nicky Rogge, Tommaso Agasisti, Kristof De Witte. Big data and them measurement of public organizations' performance and efficiency: The state-of-the-art. *Public Policy and Administration* 0(0) 1–19 (2017). <https://doi.org/10.1177/0952076716687355>
27. Ismagilova, Elvira, Hughes, Laurie, Rana, Nripendra P., Dwivedi, Yogesh K.. Security, Privacy and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-020-10044-1>
28. Loni Hagen, Hye Seon Yi, Siana Pietri and Thomas E Keller. 2019. Processes, Potential Benefits, and Limitations of Big Data Analytics: A Case Analysis of 311 Data from C ity of Miami. In *Proceedings of dg.o 2019: 20th Annual International Conference on Digital Government Research (dg.o 2019)*, June 18, 2019, Dubai, United Arab Emirates . ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3325112.3325212>
29. Karl Lofgren and C William R Webster. The value of Big Data in government: The case of 'smart cities'. *Big Data & Society* January–June 2020: 1–14. <https://doi.org/10.1177/2053951720912775>
30. Kuno Schedler, Ali Asker Guenduez and Ruth Frischknecht. How smart can government be? Exploring barriers to the adoption of smart government. *Information Polity* -1 (2019) 1–18 1. <https://doi.org/10.3233/IP-180095>
31. Muhammad Anshari & Syamimi Ariff Lim (2016): E-Government with Big Data Enabled through Smartphone for Public Services: Possibilities and Challenges, *International Journal of Public Administration*, <https://doi.org/10.1080/01900692.2016.1242619>
32. Azzone G., Big data and public policies: Opportunities and challenges. *Statistics and Probability Letters* (2018), <https://doi.org/10.1016/j.spl.2018.02.022>
33. Ali A. Guenduez, Tobias Mettler, Kuno Schedler. Technological frames in public administration: What do public managers think of big data? *Government Information Quarterly*, <https://doi.org/10.1016/j.giq.2019.101406>
34. Bram Klievink¹ & Bart-Jan Romijn¹ & Scott Cunningham¹ & Hans de Bruijn. Big data in the public sector: Uncertainties and readiness. *Inf Syst Front*. <https://doi.org/10.1007/s10796-016-9686-2>
35. Fola Malomo and Vania Sena. Data Intelligence for Local Government? Assessing the Benefits, and Barriers to Use of Big Data in the Public Sector. 1944-2866 # 2016 Policy Studies Organization. <https://doi.org/10.1002/poi3.141>
36. Anna Sexton, Elizabeth Shepherd, Oliver Duke-Williams, Alexandra Eveleigh. A balance of trust in the use of government administrative data. *Springer. Arch Sci* (2017) 17:305–330. <https://doi.org/10.1007/s10502-017-9281-4>
37. Simon Vydraa, Bram Klievink. Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*. <https://doi.org/10.1016/j.giq.2019.05.010>
38. Институт развития информационного общества. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. — 166 с
39. OECD. Digital Government Index: 2019 results. <https://dx.doi.org/10.1787/4de9f5bb-en>
40. А.В. Катин, П.С. Ершов, Ю.Е. Хохлов, С.Б. Шапошник. Модель BD4DE-ММ зрелости работы с большими данными в организации // *Информационное общество*. 2021, № 4–5. С. 259–277. https://doi.org/10.52605/16059921_2021_04_259
41. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (дата обращения 01.06.2021)

42. Форма № 3-информ. Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг. Утверждена Приказом Росстата «Об утверждении формы» от 30.07.2020 № 424.
43. Росстат. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг» за 2020 год (размещено - 13.10.2021). URL: <https://rosstat.gov.ru/storage/mediabank/3-inform.html> (дата обращения 15.10.2021)

MONITORING THE USE OF BIG DATA TECHNOLOGIES IN THE RUSSIAN PUBLIC ADMINISTRATION SYSTEM

Katin, Alexander Vladimirovich

Institute of the Information Society, head of Directorate of sectoral programs

Plekhanov Russian University of Economics, IIS-based Digital economy department, senior lecturer

Moscow, Russian Federation

alexander.katin@iis.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-Based Digital Economy Department, scientific advisor

Moscow, Russian Federation

yuri.hohlov@iis.ru

Abstract

The issues of monitoring and assessing the level of big data technologies use in the Russian system of public administration have been investigated. To develop a methodology and a conceptual monitoring scheme, as well as to select the appropriate indicators, an analysis of relevant scientific papers indexed by the Web of Science platform over the past 5 years, as well as reports of leading Russian and international organizations engaged in research on monitoring the use of digital technologies in the public administration system was carried out. The results of the analysis showed the lack of unified, universally recognized approaches to assessing and monitoring the use of technologies for working with big data in public authorities. At the same time, the topic is quite relevant, which served as an impetus for carrying out this work.

Keywords

big data, big data technologies, public administration, government, maturity level

References

1. Gosudarstvenny`e danny`e – e`to gryaznaya neft`. Glava Mincifry` Noskov predstavil Medvedevu plan raboty` s Big Data // Internet-izdanie Tadviser. 2019. URL: [https://www.tadviser.ru/index.php/Статья:Национальная_система_управления_данными_\(НСУД\)](https://www.tadviser.ru/index.php/Статья:Национальная_система_управления_данными_(НСУД)) (accessed on 01.06.2021)
2. World Bank (2017). Big Data in Action for Government: Big Data Innovation in Public Services, Policy, and Engagement. World Bank, Washington, DC. URL: <https://openknowledge.worldbank.org/handle/10986/26391> (accessed on 01.06.2021)
3. Ernst and Young. How does digital government become better government? [Electronic resource] URL: https://www.ey.com/en_gl/government-public-sector/how-does-digital-government-become-better-government (accessed on 01.06.2021)
4. SAS Institute. Big data in government: How data and analytics power public programs. URL: https://www.sas.com/en_us/insights/articles/big-data/big-data-government.html (accessed on 01.06.2021)
5. OECD (2019), The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris, <https://doi.org/10.1787/059814a7-en>
6. OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government. <https://dx.doi.org/10.1787/f64fed2a-en>
7. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol`zovaniya texnologij raboty` s bol`shimi danny`mi // Informacionnoe obshhestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
8. Haddad A., Ameen A., Isaac O., Alrajawy I., Al-Shbami A., Midhun Chakkaravarthy D. (2020) The Impact of Technology Readiness on the Big Data Adoption Among UAE Organisations. In: Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation.

- Advances in Intelligent Systems and Computing, vol 1016. Springer, Singapore.
https://doi.org/10.1007/978-981-13-9364-8_19
9. Aras Okuyucu, Nilay Yavuz. Big data maturity models for the public sector: a review of state and organizational level models. Transforming Government: People, Process and Policy © Emerald Publishing Limited 1750-6166 <https://doi.org/10.1108/TG-09-2019-0085>
 10. Hee Yeong Kim, June-Suh Cho. Data Governance Framework for Big Data Implementation with a Case of Korea. IEEE Computer Society. 978-1-5386-1996-4/17 \$31.00 © 2017 IEEE.
<https://doi.org/10.1109/BigDataCongress.2017.56>
 11. Ku, M. and Gil-Garcia, J. R. 2018. Ready for Data Analytics? Data Collection and Creation in Local Governments, In Proceedings of the 19th Annual International Conference on Digital Government Research, May 30 – June 1, Delft, the Netherlands
<https://doi.org/10.1145/3209281.3209381>
 12. Feldiansyah Bakri Nasution, Dr. Nor Erne Nazira Bazin. E-Government Maturity Model to Support System Dynamics in Public Policymaking. Proceeding of EECISI 2018, Malang - Indonesia, 16-18 Oct 2018 (p 465-471). <https://doi.org/10.1109/EECSI.2018.8752921>
 13. Manoj A. Thomasa, Joseph Cipollab, Bob Lambertc, Lemuria Carterd. Data management maturity assessment of public sector agencies. Government Information Quarterly,
<https://doi.org/10.1016/j.giq.2019.101401>
 14. Katarina Tomičić-Pupek, Igor Pihir, Martina Tomičić Furjan. Smart city initiatives in the context of digital transformation – scope, services and technologies. Journal of Contemporary Management Issues. <https://doi.org/10.30924/mjcmi.24.1.3>
 15. Tobias Brandt , Sebastian Wagner , Dirk Neumann , Prescriptive Analytics in Public-Sector Decision-Making: A Framework and Insights from Charging Infrastructure Planning, European Journal of Operational Research (2020), <https://doi.org/10.1016/j.ejor.2020.09.034>
 16. Ji Yeon Cho, Bong Gyou Lee. Creating value using public big data: comparison of driving factors from the provider’s perspective. Information Technology & People © Emerald Publishing Limited 0959-3845. <https://doi.org/10.1108/ITP-04-2019-0169>
 17. Aakash Puri. Application and Uses of Big Data Predictive Analysis in Public Sectors: A Systematic Review. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (p 539 - 543). <https://doi.org/10.1109/CTEMS.2018.8769196>
 18. Sounman Hong , Sun Hyoung Kim, Youngrok Kim and Jeongin Park. Big Data and government: Evidence of the role of Big Data for smart cities. Big Data & Society January–June 2019: 1–11.
<https://doi.org/10.1177/2053951719842543>
 19. Esmat A. Wahdain, Ahmad Suhaimi Baharudin, and Mohammad Nazir Ahmad. Big Data Analytics in the Malaysian Public Sector: The Determinants of Value Creation. Springer Nature Switzerland AG 2019 F. Saeed et al. (Eds.): IRICT 2018, AISC 843, pp. 139–150, 2019.DOI:
https://doi.org/10.1007/978-3-319-99007-1_14
 20. Kevin C. Desouza and Benoy Jacob. Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration & Society 1-22. <https://doi.org/10.1177/0095399714555751>
 21. Ismagilova E, Hughes L, Dwivedi YK et al (2019) Smart cities: Advances in research – An information systems perspective., International Journal of Information Management. 47: 88-100.
<https://doi.org/10.1016/j.ijinfomgt.2019.01.004>
 22. Mariusz Maciejewski. To do more, better, faster and more cheaply: using big data in public administration. International Review of Administrative Sciences 0(0) 1–16 (2016).
<https://doi.org/10.1177/0020852316640058>
 23. Mohammad I. Merhi, Klajdi Bregu. Effective and efficient usage of big data analytics in public sector. Transforming Government: People, Process and Policy © Emerald Publishing Limited 1750-6166. <https://doi.org/10.1108/TG-08-2019-0083>
 24. Liu Naili and Ma Lei. Research and Application of Big Data in E-government. IOP Conf. Series: Materials Science and Engineering 563 (2019) 052086. IOP Publishing.
<https://doi.org/10.1088/1757-899X/563/5/052086>
 25. Kamran Soomro , Nasir Bhutta y, Zaheer Khanz, Muhammad Atif Tahir. Smart City Big Data Analytics. An Advanced Review. Wires data mining and knowledge discovery. Volume 9, Issue 5 September/October 2019. <https://doi.org/10.1002/widm.1319>
 26. Nicky Rogge, Tommaso Agasisti, Kristof De Witte. Big data and them measurement of public organizations’ performance and efficiency: The state-of-the-art. Public Policy and Administration 0(0) 1–19 (2017). <https://doi.org/10.1177/0952076716687355>

27. Ismagilova, Elvira, Hughes, Laurie, Rana, Nripendra P., Dwivedi, Yogesh K. Security, Privacy and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-020-10044-1>
28. Loni Hagen, Hye Seon Yi, Siana Pietri and Thomas E Keller. 2019. Processes, Potential Benefits, and Limitations of Big Data Analytics: A Case Analysis of 311 Data from City of Miami. In *Proceedings of dg.o 2019: 20th Annual International Conference on Digital Government Research (dg.o 2019)*, June 18, 2019, Dubai, United Arab Emirates . ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3325112.3325212>
29. Karl Lofgren and C William R Webster. The value of Big Data in government: The case of ‘smart cities’. *Big Data & Society* January–June 2020: 1–14. <https://doi.org/10.1177/2053951720912775>
30. Kuno Schedler, Ali Asker Guenduez and Ruth Frischknecht. How smart can government be? Exploring barriers to the adoption of smart government. *Information Polity* -1 (2019) 1–18 1. <https://doi.org/10.3233/IP-180095>
31. Muhammad Anshari & Syamimi Ariff Lim (2016): E-Government with Big Data Enabled through Smartphone for Public Services: Possibilities and Challenges, *International Journal of Public Administration*, <https://doi.org/10.1080/01900692.2016.1242619>
32. Azzone G., Big data and public policies: Opportunities and challenges. *Statistics and Probability Letters* (2018), <https://doi.org/10.1016/j.spl.2018.02.022>
33. Ali A. Guenduez, Tobias Mettler, Kuno Schedler. Technological frames in public administration: What do public managers think of big data? *Government Information Quarterly*, <https://doi.org/10.1016/j.giq.2019.101406>
34. Bram Klievink¹ & Bart-Jan Romijn¹ & Scott Cunningham¹ & Hans de Bruijn. Big data in the public sector: Uncertainties and readiness. *Inf Syst Front*. <https://doi.org/10.1007/s10796-016-9686-2>
35. Fola Malomo and Vania Sena. Data Intelligence for Local Government? Assessing the Benefits, and Barriers to Use of Big Data in the Public Sector. 1944-2866 # 2016 Policy Studies Organization. <https://doi.org/10.1002/poi3.141>
36. Anna Sexton, Elizabeth Shepherd, Oliver Duke-Williams, Alexandra Eveleigh. A balance of trust in the use of government administrative data. *Springer. Arch Sci* (2017) 17:305–330. <https://doi.org/10.1007/s10502-017-9281-4>
37. Simon Vydraa, Bram Klievink. Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*, <https://doi.org/10.1016/j.giq.2019.05.010>
38. Institut razvitiya informacionnogo obshhestva. Analiz tekushhego sostoyaniya razvitiya cifrovoj e`konomiki v Rossii. M.: Institut razvitiya informacionnogo obshhestva, 2018. — 166 s.
39. OECD. Digital Government Index: 2019 results. <https://dx.doi.org/10.1787/4de9f5bb-en>
40. A.V. Katin, P.S. Ershov, Yu.E. Xoxlov, S.B. Shaposhnik. Model` BD4DE-MM zrelosti raboty` s bol`shimi dannymi v organizacii // *Informacionnoe obshhestvo*. 2021, № 4–5. S. 259–277. https://doi.org/10.52605/16059921_2021_04_259
41. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (date of application: 01.06.2021)
42. Forma № 3-inform. Svedeniya ob ispol`zovanii cifrovyy`x texnologij i proizvodstve svyazanny`x s nimi tovarov i uslug. Utverzhdena Priказom Rosstata «Ob utverzhdenii formy`» ot 30.07.2020 № 424.
43. Rosstat. Itogi federal`nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol`zovanii cifrovyy`x texnologij i proizvodstve svyazanny`x s nimi tovarov i uslug» za 2020 god (razmeshheno - 13.10.2021) (date of application: 01.06.2021)

Образование в информационном обществе**МОНИТОРИНГ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ РАБОТЫ
С БОЛЬШИМИ ДАННЫМИ В РОССИЙСКОЙ ОБРАЗОВАТЕЛЬНОЙ
СИСТЕМЕ****Елизаров Александр Михайлович**

*Доктор физико-математических наук, профессор
Институт информационных технологий и интеллектуальных систем Казанского (Приволжского)
федерального университета, кафедра программной инженерии, профессор
Казань, Российская Федерация
atelizarov@gmail.com*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества (ИРИО), председатель совета директоров
Российский экономический университет имени Г.В. Плеханова, научный руководитель базовой кафедры
цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Для проведения регулярного мониторинга уровня использования технологий работы с большими данными в российских образовательных организациях разработаны подходы к оценке текущего уровня готовности к использованию этих технологий, степени их проникновения в сферу образования, выявлению имеющихся перспективных направлений развития, сильных и слабых сторон этого процесса, а также существующих барьеров. На основе анализа документального потока по рассматриваемой теме, выполненного с помощью платформы Web of Science, установлено, что в международной практике пока отсутствуют универсальные подходы и инструменты мониторинга использования в сфере образования технологий работы с большими данными, но актуальными являются вопросы оценки зрелости и готовности к работе с большими данными в сфере образования и анализа имеющихся проблем и препятствий в их использовании. Сказанное определило выбор предметных областей мониторинга и соответствующих показателей.

Ключевые слова

большие данные; использование больших данных в образовании; мониторинг использования технологий работы с большими данными в сфере образования; уровень зрелости работы с большими данными; барьеры в использовании больших данных; BD4DE; Big Data for Digital Economy

Введение

В последние годы цифровая трансформация стала общемировым трендом, который охватывает практически все сферы деятельности, в том числе образование. Российский образовательный сектор активно использует цифровую трансформацию с целями: сделать учебный процесс позволяющим быстрее и проще обмениваться знаниями по всему миру; помочь и преподавателям, и обучаемым иметь актуальные и точные информацию и знания, необходимые для эффективного преподавания и обучения; улучшить результаты как традиционной, так и дистанционной форм обучения. В соответствии с Указом Президента РФ от 07.05.2018 № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года» [1] в качестве одной из национальных целей определено создание современной и безопасной цифровой образовательной среды. Вопросы цифровой трансформации образования нашли свое отражение и

© Елизаров А.М., Хохлов Ю.Е., 2021.

Производство и хостинг журнала «Информационное общество» осуществляются Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_166

в национальных проектах «Образование» и «Наука» [2]. В соответствии с поручением Президента России о корректировке национальных проектов от 13 июля 2020 года [3] при корректировке национального проекта «Наука» особое внимание будет уделено расширению возможных форм интеграции научных организаций, университетов и их партнеров, чтобы обеспечить новые возможности внедрения современных технологий работы с большими данными в образовании. Вместе с тем, несмотря на значительное количество проектов в сфере образования, связанных с цифровой трансформацией, реализованных в России ранее и реализуемых сегодня, технологиям работы с большими данными в них уделяется недостаточно внимания. В результате использование таких технологий, реализующих современные формы и методы обучения, в сфере российского образования остается эпизодическим.

Как отмечено в докладе [4] (с. 102), где впервые был представлен анализ текущего состояния развития цифровой экономики в России, цифровая трансформация российской сферы образования происходит постепенно по мере формирования соответствующей нормативной правовой и научно-методической базы, а общий уровень цифрового образования был оценен как формирующийся (значение показателя по 5-балльной шкале равно двум). Таким образом, сегодня Россия в направлении цифровой трансформации сферы образования отстает от ряда развитых стран, в том числе, в области внедрения технологий работы с большими данными. Вместе с тем эти технологии вывели на новый уровень использование информационно-коммуникационных технологий (ИКТ) в сфере образования – они нужны и обучающимся, и преподавателям, а также администрации образовательных учреждений, осуществляющей управление ими.

Большие данные помогают обработать опыт тысяч преподавателей и студентов и на основе его анализа сформировать новые эффективные методики преподавания. Если традиционно преподавательская методика создавалась на основе персонального опыта одного или нескольких преподавателей, то на основе больших данных такая методика становится продуктом массового опыта: можно создавать методики, адаптированные к большому количеству обучаемых; персонализировать контент; подбирать соответствующий режим обучения.

Образовательная аналитика на основе больших данных изменяет представление о формате образовательных программ и их содержании: образовательная программа превращается из утверждаемого и неизменного текста в контент, максимально нацеленный на эффективное онлайн-обучение и динамически изменяющийся в результате анализа данных и взаимодействия с обучающимися. В итоге создаются так называемые «умные программы» и «умные учебные планы», а программы учебных курсов становятся метапредметными. Кроме того, большие данные позволяют эффективнее вести мониторинг успеваемости и персонализировать программы обучения.

Еще одно важное направление использования технологий работы с большими данными связано с их применением в управлении образованием – сложном и многогранном процессе, ориентированном на оптимальную и продуктивную организацию функционирования образования в нашей стране. Отметим, что в соответствии со ст. 89 Федерального закона «Об образовании в Российской Федерации» [5] управление образованием осуществляется на трех уровнях: федеральном, региональном и местном (муниципальном). Каждый из этих уровней имеет свои особенности в использовании технологий работы с большими данными.

Названные выше, а также многие другие возможности, направления и примеры использования больших данных в образовании обсуждались в ряде зарубежных публикаций, начиная с 2010-х годов (например, [6–8]). Немного позднее появились обзорные работы и книги, содержащие соответствующую информацию в обобщенном виде (см., например, [9–11] и библиографию в них). Интересен также опыт использования больших данных в образовательных системах отдельных зарубежных стран (например, [12–14]). Имеются некоторые результаты применения больших данных в российской образовательной системе (см. [15–17]).

Чтобы оценить текущий уровень использования технологий работы с большими данными в российских образовательных организациях, степень проникновения этих технологий в сферу образования, выявить имеющиеся перспективные направления развития, сильные и слабые стороны этого процесса, а также существующие барьеры, необходимо регулярно проводить мониторинг уровня использования технологий работы с большими данными в этих организациях.

Настоящее исследование опирается на подходы к мониторингу и оценке производства, использования и воздействия технологий, решений и услуг работы с большими данными [18],

разработанные в проекте «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» Национальной технологической инициативы. Применению разработанных подходов к организации такого мониторинга в организациях высшего и дополнительного профессионального образования как одного из ключевых компонентов концептуальной схемы мониторинга BD4DE (Big Data for Digital Economy) и посвящена настоящая работа.

1 Предметная область мониторинга и обзор литературы

Предметной областью мониторинга является деятельность российских организаций (образовательных или осуществляющих обучение) в части использования ими технологий работы с большими данными. Всюду в дальнейшем для краткости будем называть организации, образовательные или осуществляющие обучение, просто образовательными организациями.

Отметим, что в соответствии со ст. 89 Федерального закона «Об образовании в Российской Федерации» [5] образовательная организация – это некоммерческая организация, осуществляющая на основании соответствующей лицензии образовательную деятельность в качестве основного вида деятельности в соответствии с целями, ради достижения которых такая организация создана; а организация, осуществляющая обучение, – это юридическое лицо, осуществляющее на основании полученной лицензии наряду с основной своей деятельностью образовательную деятельность в качестве дополнительного вида деятельности (такowymi, например, являются институты Российской академии наук). Отметим, что названный закон сформировал правовое поле цифровой трансформации системы образования. В соответствии с этим законом сегодня в России реализуются: общее образование (уровни – дошкольное, начальное, основное и среднее) и профессиональное образование (уровни – среднее профессиональное, высшее образование – бакалавриат; высшее образование – специалитет, магистратура; высшее образование – подготовка кадров высшей квалификации).

На сегодняшний день говорить о внедрении технологий работы с большими данными в российской сфере образования имеет смысл, прежде всего, для организаций высшего профессионального и дополнительного профессионального образования, а также научно-исследовательских организаций, осуществляющих подготовку научно-педагогических кадров в аспирантуре. Именно на эти организации нацелен организуемый мониторинг использования технологий работы с большими данными – для организаций, осуществляющих образовательную деятельность по образовательным программам начального, основного и среднего общего образования, сегодня актуальны лишь базовые вопросы цифровизации своей деятельности.

По данным [19], на конец 2020 года организаций, осуществляющих образовательную деятельность по программам высшего образования было 705 (из них частных – 213); организаций высшего образования и научных организаций, осуществляющих образовательную деятельность по программам бакалавриата, специалитета и магистратуры, – 724 (из них частных – 229); организаций, осуществляющих подготовку научно-педагогических кадров в аспирантуре – 1187 (из них научно-исследовательских – 599, организаций высшего образования – 567, организаций дополнительного профессионального образования – 17). Анализ статистических данных за указанный и предыдущие годы показал, что в последнее время в России численность вузов продолжает сокращаться, при этом количество частных вузов, которые, по-видимому, не выдерживают конкуренции с государственными организациями, убывает быстрее.

Оценка глобального российского рынка больших данных в сфере образования затруднительна. Вместе с тем имеются многочисленные данные исследований рынка онлайн-образования, существенной составляющей частью которого является рынок технологий работы с большими данными. По информации отчета «Исследование российского рынка онлайн-образования» [20], мировой рынок онлайн-образования к 2023 году должен достичь 282 млрд долларов США, в то время как в 2017 году он измерялся суммой 159 млрд долларов США, в 2018-м – 190 млрд долларов США, в 2019-м – 205 млрд долларов США. Следовательно, среднегодовой темп прироста этого рынка в указанные годы составил 7–10%. Не меньшим можно прогнозировать и ежегодный прирост рынка технологий работы с большими данными в сфере образования.

Все имеющиеся и вновь возникающие барьеры в использовании технологий работы с большими данными характерны для всех сфер возможного их внедрения и имеют свою историю возникновения и преодоления. Отметим в этой связи, что еще в 2014 году Accenture [21], ведущей

международной компанией в области стратегии, консалтинга, цифровизации бизнеса, технологий и операций, были даны следующие оценки основных проблем при внедрении проектов больших данных (указаны % исследованных организаций, столкнувшихся с названными проблемами): вопросы безопасности – 51%; ограниченный бюджет – 47%; нехватка персонала для внедрения технологий больших данных – 41%; нехватка персонала для ведения проектов работы с большими данными и аналитики – 37%; сложности интеграции с существующими системами – 35%; ограниченное число поставщиков данных – 33%; неготовность к работе с большими данными – 27%.

По сведениям журнала Computerworld Россия [22], распределение ключевых проблем в рамках отдельного проекта внедрения больших данных в любой сфере деятельности усредненно выглядит так: сбор данных и их качество – 30%; недостаток финансирования – 19%; отсутствие заинтересованности – 19%; необходимость модернизации инфраструктуры – 11%; нехватка компетенций – 7%, безопасность – 7%; другое – 7%.

В стратегии развития рынка больших данных на ближайшее пятилетие, разработанной Ассоциацией больших данных [23], отмечено, что России для раскрытия потенциала больших данных необходимо преодолеть следующие барьеры:

- нехватку специалистов (нужно перестроить систему высшего образования);
- ограниченное предложение недорогих масштабируемых ресурсов хранения и обработки данных;
- ограниченную способность получать доступ, обрабатывать и обмениваться различными типами данных (на рынке не сформировалась экосистема обмена данными и основных вариантов ее использования);
- сложности в проведении исследований из-за отсутствия необходимых среды и ресурсов;
- отсутствие широкомасштабного внедрения больших данных в традиционных отраслях и целевой поддержки продуктов на основе больших данных в инновационных проектах.

В целом указанные проблемы и барьеры характерны и для сферы образования при внедрении в нее технологий работы с большими данными. Тем не менее, можно выделить следующие основные барьеры, характерные именно для образовательных организаций.

Прежде всего, это ограниченность выделяемых финансовых ресурсов и относительно небольшая величина рынка платных образовательных услуг в России по сравнению со средствами бюджета, выделяемыми на эту сферу деятельности. Поэтому в рамках мониторинга необходимо оценить долю образовательных организаций, считающих, что затраты на технологии работы с большими данными слишком высоки по сравнению с выгодами, которые потенциально могут быть получены.

Серьезным препятствием для развития цифрового образования является нехватка квалифицированных кадров (как среди основного преподавательского, так и вспомогательного и административного персонала). Она характерна для образовательных организаций в силу неконкурентного уровня заработной платы их сотрудников по сравнению с частным сектором. В рамках мониторинга барьеров использования технологий работы с большими данными в образовательных организациях важно оценить долю названных организаций, считающих, что у них недостаточно человеческих ресурсов, знаний и навыков.

Критическим барьером для полноценного внедрения и использования технологий работы с большими данными в образовательных организациях является отсутствие в них современной цифровой инфраструктуры.

Все показатели, характеризующие барьеры в использовании технологий работы с большими данными в образовательных организациях, названные выше, включены в перечень показателей мониторинга, приведенных в следующем разделе.

Применение технологий, решений и услуг работы с большими данными является относительно новой деятельностью для российских образовательных организаций. Несмотря на значительный потенциал этого класса цифровых технологий, на текущий момент времени они не являются приоритетными. Для мониторинга уровня готовности названных организаций к работе с большими данными необходимо проводить оценку доли среди таких организаций тех, которые считают, что эти технологии не являются приоритетом для их деятельности, оценивают недостаточными источники больших данных для решения текущих задач, а также полагают, что использование больших данных бесполезно.

Для оценки имеющегося мирового документального потока по использованию технологий работы с большими данными в образовании была использована платформа Web of Science (WoS), содержащая одну из самых полных и авторитетных междисциплинарных библиографических баз данных о научных исследованиях и инструменты для анализа сведений о публикациях и цитированиях. На первом этапе был сформирован исходный список ключевых слов, характеризующих предметную область технологий работы с большими данными, при этом основными критериями отбора терминов в качестве ключевых слов были (см. [18]):

- специфичность – термин используется в рамках области технологий работы с большими данными;
- полезность для целей исследования – термин расширяет искомый массив и позволяет идентифицировать публикации, которые не имеют в названии или других метаданных термина «big data»;
- отсутствие «шума» в результатах поиска – т. е. основная часть публикаций, содержащих предлагаемый термин в названии, ключевых словах или аннотации, должна относиться к указанной области исследований и разработок.

На втором этапе в ходе независимого экспертного опроса исходный список ключевых слов был верифицирован – был сформирован поисковый запрос для отбора публикаций в области технологий работы с большими данными, учитывающий различные варианты написания терминов. Этот запрос был использован в поисковом поле «Тема» (TS – поиск по названиям, ключевым словам и аннотациям) для формирования в WoS анализируемого массива публикаций за 2016–2020 гг.

Отметим, что первоначальный выбор более широкого поискового запроса для предметной области, связанной с использованием технологий работы с большими данными, был продиктован тем, что, хотя эти технологии и представляют собой частный случай информационно-коммуникационных технологий (ИКТ), терминология и классификация (суб)технологий больших данных окончательно не сформировались. Поэтому, чтобы не потерять релевантные публикации, было принято решение использовать достаточно общие термины.

Дальнейшее сужение полученного массива проводилось путем выделения подмножества публикаций, связанных с использованием больших данных в образовании. Для этого вновь методом экспертных оценок с последующей верификацией был сформулирован следующий поисковый запрос, также учитывающий различные варианты написания терминов:

“Big data*” OR “Bigdata” OR “Large dataset*” AND “education*” OR “learning” OR “higher education” OR “educational big data” OR “educational data mining” OR “learning analytics” OR “Big Data analytics in higher education” OR “management of the educational process*” OR “individual educational trajectory” OR “e-learning” OR “digital learning”

Результат применения приведенного запроса привел к выделению 12857 научных публикаций, причем добавление к запросу AND “monitoring use in education” почти полностью обнулило количество публикаций в выделенном множестве, что свидетельствует о малом представительстве тематики, связанной с мониторингом. Поэтому было принято решение сосредоточиться только на обзорных статьях, которых в изначально выделенном поисковом массиве оказалось 861. Далее был проведен анализ тех из них, которые имеют прямое отношение к изучаемой предметной области. Таковых за последнее пятилетие оказалось около 20. Некоторые из них уже процитированы выше (см. [6–17]). Обзорные материалы по использованию технологий работы с большими данными, ориентированные на систему высшего образования, содержатся также в работах [24–37], в том числе нацеленные на управление образованием – в [38]. Использование обзорных работ позволило в достаточно полной мере охватить все имеющиеся публикации по интересующей нас тематике. Остановимся дополнительно на работах, содержащих некоторые результаты библиометрического анализа.

Прежде всего, выделим работу [39], в которой в результате проведенного библиометрического исследования по тематике использования больших данных в образовании с использованием баз данных WoS, Scopus, ERIC и PsycINFO за 2010–2019 годы выделен 1491 документ, дан анализ полученных результатов, сформированы конфигурации определенных журналов, стран и авторов по рассматриваемой тематике и оценены перспективы будущих исследований.

В работе [40] в результате выполненного обзора журнальных статей по тематике работы с большими данными в образовании, опубликованных с 2010 по 2018 годы, было отобрано 143 статьи, которые были разделены на группы в соответствии с четырьмя следующими основными выделенными проблемами, возникающими при использовании больших данных: отсутствие моделирования поведения обучаемых и нерациональная трата ресурсов; несоответствие учебных программ и стратегий обучения; обеспечение гарантий качества обучения; нерешенные вопросы обеспечения конфиденциальности и этики. В качестве наиболее подходящего для обеспечения качества обучения и потенциальных решений в случае отсутствия моделирования поведения обучающихся и нерационального использования ресурсов назван метод интеллектуального анализа данных.

Анализ описанного выше документального потока показал, что на текущий момент времени в международной практике отсутствуют универсальные подходы и инструменты мониторинга использования цифровых технологий в сфере образования, в том числе технологий работы с большими данными, а имеющиеся статистические данные ОЭСР и других всемирных организаций не увязаны напрямую с использованием ИКТ в образовании. Вместе с тем, в большинстве обзорных работ, упомянутых выше, в той или иной форме говорится о зрелости и готовности к работе с большими данными в сфере образования, а также об имеющихся проблемах и препятствиях в их использовании. Сказанное определило выбор предметных областей мониторинга и соответствующих показателей.

2 Концептуальная схема и показатели мониторинга

Под концептуальной схемой понимается совокупность предметных областей мониторинга и оценки, структура и взаимосвязи которых определяются научными представлениями об области мониторинга, а также его целями и задачами.

Разработанная концептуальная схема мониторинга и оценки уровня использования технологий работы с большими данными в сфере образования включает три компонента и представлена на рис. 1.



Рис. 1. Концептуальная схема мониторинга уровня использования технологий работы с большими данными в сфере образования

Подход к оценке использования технологий работы с большими данными в сфере образования опирается на названную концептуальную схему и включает три предметные области мониторинга: оценка уровня зрелости работы с большими данными в образовательных

организациях, использование технологий работы с большими данными в образовании, а также барьеры в использовании этих технологий образовательными организациями. Данный подход позволяет получить всестороннее представление о потенциальных точках роста и выделить проблемные зоны распространения технологий работы с большими данными в образовательных организациях, а также выявить специфику их применения в отдельных организациях этого сектора деятельности.

2.1 Показатели уровня зрелости работы с большими данными

Для оценки уровня зрелости работы с большими данными в организациях сферы образования должны учитываться специфика деловых процессов в этих организациях, а также цели использования технологий работы с большими данными. В основе оценки лежит модель зрелости BD4DE-MM работы с большими данными в организациях, разработанная и подробно описанная в работе [41]. Эта модель основана на ведущих мировых методиках проведения подобных исследований и включает следующие семь размерностей:

1. стратегия и регулирование;
2. кадры и лидерство;
3. данные;
4. инструменты и аналитика;
5. инфраструктура и безопасность;
6. организация работы;
7. воздействие.

Для оценки используется пятибалльная порядковая шкала, задающая следующие уровни зрелости.

1. Начальный уровень — организация не использует большие данные.
2. Формирующийся уровень — организация, начинающая использовать большие данные.
3. Средний уровень — организация, использующая большие данные.
4. Продвинутый уровень — организация, масштабно использующая зрелые технологии работы с большими данными.
5. Высокий уровень — организация-лидер, масштабно использующая большие данные и внедряющая новые технологии работы с большими данными.

Для отдельной образовательной организации возможность проведения (само)оценки уровня зрелости работы с большими данными в соответствии с приведенными показателями позволит выявить имеющиеся сильные и слабые стороны этой деятельности, повысить эффективность и сравнить себя со схожими организациями.

Наряду с этим, не менее важной является оценка уровня зрелости работы с большими данными во всей сфере образования в целом, позволяющая сравнить состояние дел с другими сферами деятельности. Так как воздействие (социальные и экономические эффекты) проявляется лишь на продвинутом уровне зрелости работы с большими данными в организациях, то в качестве меры зрелости работы с большими данными для всей сферы образования будем использовать долю организаций, которые достигли продвинутого или высокого уровня зрелости, среди всех организаций данной сферы.

Подобные интегральные показатели дают возможность сравнивать отрасли экономики или сферы деятельности между собой, аналогичным образом могут быть построены измеримые показатели для оценки отдельных аспектов использования больших данных. Например, сопоставлять кадровое обеспечение работы с большими данными или совместное использование больших данных в различных деловых процессах организаций рассматриваемой сферы деятельности с другими. Регулярный мониторинг и интегральная оценка уровня зрелости дают возможность более рационально и обоснованно выстраивать государственную политику в направлении развития технологий работы с большими данными и внедрения их в сферу образования.

В настоящем исследовании использован интегральный показатель:

- (ЦЗО-01) Доля организаций, имеющих продвинутый и выше уровень зрелости работы с большими данными, среди всех организаций сферы образования,

который характеризует уровень зрелости работы с большими данными для всей сферы образования.

Интегральный показатель ЦЗО-01 представляет собой среднее арифметическое 7 отдельных показателей, соответствующих размерностям модели зрелости BBD4DE, и вычисляется как доля организаций, которые обладают соответствующим свойством, среди всех организаций сферы образования.

- (ЦЗО-01-01) Доля организаций, имеющих продвинутый и выше уровень зрелости стратегии и регулирования работы с большими данными.

Этот показатель характеризует отношение в организациях сферы образования к стратегическому планированию и регулированию использования технологий работы с большими данными в своей деятельности. При этом стратегическое планирование не сводится к наличию в организации стратегии использования больших данных в виде отдельного документа или раздела стратегии цифровой трансформации, она должна соответствовать общей стратегии развития организации и быть подкреплена планом ее реализации, обеспеченным необходимыми ресурсами.

- (ЦЗО-01-02) Доля организаций, имеющих продвинутый и выше уровень кадрового обеспечения работы с большими данными.

Уровень осведомленности, знаний и компетенций в области работы с большими данными как профессорско-преподавательского состава, так и обучающихся играет ключевую роль в их успешном использовании. Еще одним ключевым фактором успеха – наличие в организации руководителей, заинтересованных во внедрении технологий работы с большими данными.

- (ЦЗО-01-03) Доля организаций, имеющих продвинутый и выше уровень обеспеченности качественными большими данными.

Обеспеченность деловых процессов большими данными, их качество и совместное использование позволяют существенно повысить эффективность функционирования организации.

- (ЦЗО-01-04) Доля организаций, имеющих продвинутый и выше уровень использования аналитики и инструментов работы с большими данными.

Регулярность использования аналитики больших данных и других специализированных инструментов в деловых процессах образовательной организации (прежде всего – в образовательном процессе) также является предпосылкой для достижения экономических или социальных эффектов.

- (ЦЗО-01-05) Доля организаций, имеющих продвинутый и выше уровень зрелости цифровой инфраструктуры и безопасности для работы с большими данными.

Наличие цифровой инфраструктуры (как внутренней, так и внешней, например, облачной), которая является безопасной и масштабируемой, представляет собой необходимое условие для успешной работы с большими данными в организации.

- (ЦЗО-01-06) Доля организаций, имеющих продвинутый и выше уровень зрелости организационных механизмов работы с большими данными.

При оценке системы (стратегического) управления работы с большими данными в организации учитывается как наличие необходимых организационных единиц (например, ответственного руководителя, профильного подразделения, центра компетенций), так и уровень регламентации управленческих процессов для работы с большими данными.

Как интегральный показатель ЦЗО-01 зрелости работы с большими данными в образовательной сфере, так и отдельные показатели ЦЗО-01-01 – ЦЗО-01-06 не являются уникальными для анализируемой сферы деятельности и могут использоваться (с учетом специфики соответствующих процессов) в других сферах, что дает возможности сравнения. В то же время, применительно к сфере образования наибольший интерес представляет использование больших данных в образовательных процессах и управлении образованием.

2.2 Показатели использования технологий работы с большими данными

Для мониторинга использования технологий работы с большими данными в образовательных организациях отобраны показатели, которые позволяют оценить уровень проникновения указанных технологий в деятельность всех образовательных организаций, включая управление образовательным процессом. Большинство показателей этого компонента концептуальной схемы

мониторинга BD4DE позволяет проводить международные сопоставления, т. к. они используются в рамках опроса организаций об использовании ИКТ в Европейском союзе (см. [42]). При этом все показатели вычисляются как доля организаций, которые обладают соответствующим свойством, среди всех организаций сферы образования.

- (ИБДО-01) Доля образовательных организаций, использующих технологии работы с большими данными

Такой показатель позволит оценить уровень использования этих технологий в образовательных организациях на различных стадиях жизненного цикла работы с данными, а именно: при генерации больших данных, их сборе, обработке, хранении, анализе и визуализации.

Для оценки уровня компетенций, а также заделов и возможностей, имеющихся у образовательных организаций, используется показатель:

- (ИБДО-02) Доля образовательных организаций, выполняющих работу с большими данными собственными силами или силами внешних компаний.

Сфера образования генерирует значительный объем данных разного формата и из множества источников. Все эти данные могут быть полезны как компаниям, которые осуществляют разработки в сфере работы с большими данными, так и самим образовательным организациям, которые используют их для повышения качества оказываемых услуг. Чтобы оценить, насколько эти данные востребованы при оценке уровня использования технологий работы с большими данными в образовательных организациях, применяются следующие показатели, позволяющие оценить долю образовательных организаций, предоставляющих (продающих), а также получающих (покупающих) доступ к большим данным:

- (ИБДО-03) Доля образовательных организаций, предоставляющих (продающих), а также получающих (покупающих) доступ к большим данным;
- (ИБДО-04) Доля образовательных организаций, использующих большие данные для повышения качества оказываемых услуг.

В России выделяются значительные бюджетные средства на цифровизацию сферы образования. Однако стоит отметить, что выделить из общего бюджета на внедрение цифровых технологий в образовательных организациях затраты на внедрение технологий работы с большими данными не всегда представляется возможным. Поэтому в рамках проводимого мониторинга используется показатель

- (ИБДО-05) Доля затрат образовательных организаций на работу с большими данными в общих затратах на цифровые технологии,

который дает примерную оценку рынка технологий работы с большими данными в российской системе образования.

Одним из самых действенных инструментов использования больших данных является осуществление аналитики. В сфере образования аналитика больших данных может применяться для повышения качества образовательных услуг либо эффективности управления образовательным учреждением. Поэтому одним из важнейших направлений настоящего мониторинга является определение организаций сферы образования, использующих аналитику больших данных в тех или иных целях:

- (ИБДО-06) Доля образовательных организаций, применяющих аналитику больших данных.

При расчете этого показателя представляет интерес не только определить число организаций образовательной сферы, которые вообще используют аналитику больших данных, но и оценить такие направления использования, как маркетинг, образовательный процесс, административные процессы, управление кадрами.

Отметим, что показатели ИБДО-01 – ИБДО-06 оценки уровня использования технологий работы с большими данными образовательными организациями аналогичным образом могут использоваться и в других сферах деятельности (см. [43–46]), они лишь уточнены в части целеполагания использования.

Мониторинг использования технологий работы с большими данными в организациях сферы образования дает возможность выявить имеющиеся приоритеты и сравнить достигнутый уровень с показателями других сфер деятельности. Наличие такой оценки, регулярно корректируемой в ходе мониторинга, поможет, в частности, компаниям – поставщикам технологий, решений и услуг для

работы с большими данными – более эффективно выстраивать свою деятельность при формировании предложений для сферы образования.

2.3 Показатели, характеризующие барьеры использования больших данных

Анализ барьеров в использовании технологий работы с большими данными в сфере образования является важным элементом мониторинга, поскольку, только минимизировав или устранив препятствия в полноценном использовании названных технологий, можно говорить о потенциальных эффектах от их использования для сферы образования в целом. Наличие обоснованной оценки причин неиспользования технологий работы с большими данными будет полезно всем сторонам, ответственным за развитие российского образования, а также бизнесу, который сможет рационально реагировать на проблемы и возможности, имеющиеся на рынке.

Как отмечают аналитики и эксперты (см., в частности, анализ документального потока в разделе 1), существуют барьеры, которые тормозят внедрение технологий работы с большими данными в образовательный процесс. К ним относятся рассредоточение данных и ограниченный доступ к ним, опасность потери данных, отсутствие обученного персонала, проблемы с использованием персональных данных и многое другое.

Показатели, включенные в систему мониторинга BD4DE и характеризующие барьеры в использовании технологий работы с большими данными образовательными организациями, согласованы с модельной анкетой Евростата (см. например, [42]), по которой проводился опрос европейских организаций в 2018–2020 годах. Это позволяет провести международные сопоставления (с показателями стран-членов ЕС). Как и ранее, показатели рассчитываются в долях организаций, которые отметили соответствующий барьер, среди всех организаций сферы образования.

- (ББДО-01) Доля образовательных организаций, считающих, что работа с большими данными не является приоритетом для их основной деятельности.

Этот показатель характеризует организации, для которых работа с большими данными не стала стратегическим приоритетом.

- (ББДО-02) Доля образовательных организаций, считающих, что затраты на работу с большими данными слишком высоки по сравнению с выгодами.

Любая технологическая инновация требует инвестиций, особенно в случае, когда быстрый возврат инвестиций для организаций не очевиден. В эту категорию попадают организации, которые не имеют достаточно средств или не готовы их тратить на внедрение технологий работы с большими данными, так как считают, что их вложения не окупятся.

- (ББДО-03) Доля образовательных организаций, считающих, что работа с большими данными для них бесполезна.

Показатель характеризует наличие организаций, которые не видят выгод от использования больших данных в своей деятельности.

- (ББДО-04) Доля образовательных организаций, считающих, что у них недостаточно человеческих ресурсов, знаний, навыков для работы с большими данными.

Наличие в организации подготовленных кадров – важный фактор, влияющий на внедрение цифровых технологий, а их отсутствие служит барьером. К этому могут добавляться и трудности в поиске и привлечении к работе в организацию специалистов с требуемыми компетенциями.

- (ББДО-05) Доля образовательных организаций, считающих, что у них недостаточно развита цифровая инфраструктура для работы с большими данными.

Технологии оперирования большими данными весьма требовательны к используемым программным и аппаратным средствам, поэтому важно понимать, каков уровень зрелости цифровой инфраструктуры в организациях сферы образования в целом, чтобы принимать адекватные меры поддержки.

- (ББДО-06) Доля образовательных организаций, считающих, что источников больших данных (как в организации, так и за ее пределами) недостаточно для решения текущих задач с помощью аналитики больших данных.

Этот показатель характеризует трудности обеспечения организации большими массивами данных из всевозможных источников, которые позволят решать текущие задачи методами анализа больших данных.

- (ББДО-07) Доля образовательных организаций, считающих, что качество имеющихся больших массивов данных (как в организации, так и за ее пределами) не позволяет решать текущие задачи.

Качество данных, в том числе больших массивов данных – это степень их пригодности к обработке и анализу, а также соответствие обязательным или специальным требованиям. Этот показатель характеризует наличие организаций, которые не могут обеспечить необходимое качество больших данных для решения текущих задач.

- (ББДО-08) Доля образовательных организаций, отмечающих трудности соблюдения законодательства о персональных данных при работе с большими данными.

Законодательство о персональных данных запрещает обработку персональных данных субъекта без его явного согласия и указания целей обработки, что существенно ограничивает возможности организаций по сбору и обработке больших массивов пользовательских данных.

- (ББДО-09) Доля образовательных организаций, считающих использование аналитики больших данных не этичным.

Современные инструменты работы с большими данными могут не только способствовать организациям в решении текущих задач, но и вторгаться в личную жизнь человека, что ставит вопрос об этичности использования технологий работы с большими данными. Такой показатель характеризует организации, которые воздерживаются от использования больших данных по названным причинам.

Вновь отметим, что показатели ББДО-01 – ББДО-09, характеризующие барьеры в использовании больших данных образовательными организациями, аналогичным образом могут применяться в других сферах деятельности (см. [43–46]), здесь они лишь уточнены в части целеполагания использования.

3 Результаты пилотной реализации мониторинга

По результатам первой стадии пилотной реализации предложенной системы мониторинга BD4DE применительно к использованию технологий работы с большими данными в сфере образования была разработана анкета для обследования образовательных организаций.

На второй стадии пилотной реализации проводилась экспертная оценка разработанной анкеты с привлечением эксперта-специалиста по применению цифровых технологий в образовании. Им были сделаны предложения по уточнению формулировок, в частности, при описании основных образовательных процессов, а также при формулировке квалификационных требований о наличии у соискателей навыков по работе с большими данными при размещении вакансий для сотрудников-неспециалистов.

Кроме того, для учета специфики сферы образования в раздел анкеты с вопросом «Для каких целей применяется аналитика больших данных в образовательной организации» было предложено внести дополнительные вопросы, отражающие такую специфику. В результате были добавлены следующие направления использования аналитики больших данных: «Повышение эффективности и результативности процессов обучения», «Кастомизация (адаптация, настройка под запросы конкретного потребителя) программ и процессов обучения» и «Прогнозирование карьеры обучающихся».

На последней стадии пилотной реализации перечисленные выше замечания и предложения были учтены в финальной версии анкеты для обследования образовательных организаций.

На основании официальных статистических данных за 2020 г. [48] были рассчитаны значения для показателей ИБДО-01 и, частично, для ИБДО-06 (рисунок 2).

Доля образовательных организаций, использовавших технологии сбора, обработки и анализа больших данных в 2020 году, составила 27,7%. В то же время выполняют анализ больших данных меньше половины образовательных организаций – 12,4 %, причем 7,5% осуществляют данную деятельность только собственными силами.

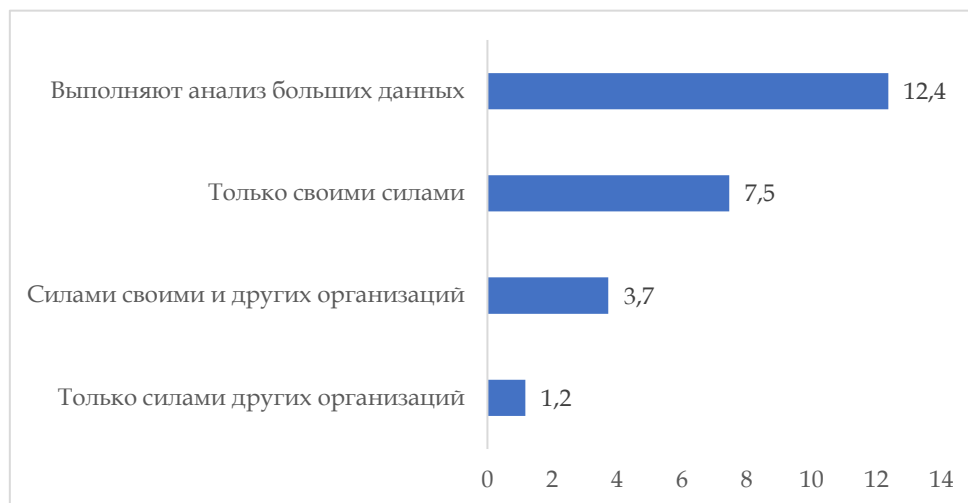


Рис. 2. Доля образовательных организаций, выполняющих анализ больших данных, 2020, %
Источник: [48], расчеты авторов

На следующем этапе исследований для полноценного мониторинга будет проводиться опрос образовательных организаций, причем некоторые показатели будут получены методом экспертных опросов.

Заключение

Подводя итог, отметим, что подходы к оценке текущего уровня готовности к использованию технологий работы с большими данными разработаны для проведения регулярного мониторинга уровня использования этих технологий в российских образовательных организациях, оценки степени их проникновения в сферу образования, выявления имеющихся перспективных направлений развития, сильных и слабых сторон этого процесса, а также существующих барьеров. Предложенный подход к исследованию интенсивности и проблем в использовании названных технологий в российских образовательных организациях высшего и дополнительного профессионального образования позволит сформировать комплексное представление о потенциальных точках роста и проблемных зонах применения технологий работы с большими данными в названной сфере.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Литература

1. Указ Президента Российской Федерации от 07.05.2018 г. № 204. О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года. URL: <http://kremlin.ru/acts/bank/43027> (дата обращения 01.06.2021).
2. Паспорт национального проекта «Образование». URL: <http://government.ru/info/35566/> (дата обращения 01.06.2021); Паспорт национального проекта «Наука». URL: <http://government.ru/info/35565/> (дата обращения 01.06.2021).

3. Заседание Совета по стратегическому развитию и национальным проектам от 13.07.2020 г. URL: <http://kremlin.ru/events/president/news/63635> (дата обращения 01.06.2021).
4. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. 166 с. URL: <https://iis.ru/wp-content/uploads/2020/12/DECARussia2018rus.pdf> (дата обращения 01.06.2021).
5. Федеральный закон от 29.12.2012 г. № 273-ФЗ «Об образовании в Российской Федерации». URL: <http://kremlin.ru/acts/bank/36698> (дата обращения 01.06.2021).
6. Samson Oluwaseun Fadiya, Serdar Saydam, Emeka Joshua Chukwuemeka. Big Data in Education; Future Technology Integration//The International Journal of Science & Technology. 2014. Vol. 2, Issue 8. P. 65-69. URL: https://www.researchgate.net/publication/281508011_Big_Data_in_Education_Future_Technology_Integration (дата обращения 01.06.2021).
7. Athanasios S. Drigas. The Use of Big Data in Education // IJCSI International Journal of Computer Science. September 2014. Vol. 11, Issue 5, No. 1. P. 58-63. URL: <https://www.ijcsi.org/papers/IJCSI-11-5-1-58-63.pdf> (дата обращения 01.06.2021).
8. Jaimin N. Undavia, Sheenal Patel, Atul Patel. Future trends and scopes of Big Data Analytics in the field of Education// International Journal of Engineering and Technology (IJET). July 2017. Vol. 9, No. 3S. P. 9-14. <https://doi.org/10.21817/ijet/2017/v9i3/170903S002>
9. Tamaro Green. Big Data in Education Technology: Improvements in International Research Design: Data Driven Testing, Analysis, and Programming Paperback. March 24, 2017. 149 p.
10. Tamaro J. Green. Big Data Analytics in Higher Education. May 2017. <https://doi.org/10.13140/RG.2.2.22565.50407>
11. C. Fischer, Z.A. Pardos, R.S. Baker et al. Mining big data in education: Affordances and challenges. Review of Research in Education. 2020. Vol. 44, No. 1. P. 130-160. <https://doi.org/10.3102/0091732X20903304>
12. M.Sh. Hajirahimova, A.S. Aliyeva. Methodologies and indicators for Big Data measurement // Современные информационные технологии и ИТ-образование. 2017. Том 13, № 2. С. 81-88. <https://doi.org/10.25559/SITITO.2017.2.239>
13. Noura A. Alsheikh. Developing an integrated framework to utilize big data for higher education institutions in Saudi Arabia // International Journal of Computer Science & Information Technology (IJCSIT). February 2019, Vol. 11, No. 1. P. 31-42. <https://doi.org/10.5121/ijcsit.2019.11103>
14. Şeyhmus Aydoğdu. Educational data mining studies in Turkey: a systematic review// Turkish Online Journal of Distance Education. 2020. Vol. 21, No. 3, Article 12. Pub Date: 2020-07-01. <https://doi.org/10.17718/tojde.762046>
15. В.В. Утёмов, П.М. Горев. Развитие образовательных систем на основе технологии Big Data // Концепт. 2018. № 6. URL: <https://cyberleninka.ru/article/n/razvitie-obrazovatelnyh-sistem-na-osnove-tehnologii-big-data> (дата обращения 01.06.2021).
16. Б.А. Кондратенко, А.Б. Кондратенко. Перспективы применения анализа больших данных в современном образовании // Вестник Балтийского федерального университета им. И. Канта. Сер.: Филология, педагогика, психология. 2018. № 1. С. 117-126. URL: https://journals.kantiana.ru/upload/iblock/d2a/%D0%9A%D0%BE%D0%BD%D0%B4%D1%80%D0%B0%D1%82%D0%B5%D0%BD%D0%BA%D0%BE%20%D0%91.%20%D0%90.,%20%D0%9A%D0%BE%D0%BD%D0%B4%D1%80%D0%B0%D1%82%D0%B5%D0%BD%D0%BA%D0%BE_117-126.pdf (дата обращения 01.06.2021).
17. Ю.В. Малышева. Технология больших данных в российском образовании // Научные исследования XXI века. 2020. Том 1. №3. С. 59-62.
18. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
19. Реестр открытых данных Минобрнауки. <https://minobrnauki.gov.ru/opendata/> (дата обращения 01.06.2021).
20. Исследование рынка онлайн-обучения 2020. URL: <http://research.edmarket.ru/> <https://minobrnauki.gov.ru/opendata/> (дата обращения 01.06.2021).

21. Международная компания в области стратегии, консалтинга, цифровизации бизнеса, технологий и операций Accenture. URL: <https://www.accenture.com/ru-ru> (дата обращения 01.06.2021).
22. Журнал Computerworld Россия. URL: <https://www.computerworld.ru/> (дата обращения 01.06.2021).
23. Ассоциация больших данных. URL: <https://rubda.ru/> (дата обращения 01.06.2021).
24. Vatsala, Rutuja Jadhav, R. Sathyaraj. A Review of Big Data Analytics in Sector of Higher Education // Int. Journal of Engineering Research and Application. 2017. Vol. 7, Issue 6 (Part 2). P. 25-32. URL: <https://doi.org/10.9790/9622-0706022532> (дата обращения 01.06.2021).
25. Big Data and Learning Analytics in Higher Education. Current Theory and Practice. Ben Kei Daniel (Ed.). Springer, 2017. 287 p. <https://doi.org/10.1007/978-3-319-06520-5>
26. Chad Laux, Na Li, Corey Seliger, John Springer. Impacting Big Data analytics in higher education through Six Sigma techniques // International Journal of Productivity and Performance Management. 2017. Vol. 66, Issue 5. P. 662-679, <https://doi.org/10.1108/IJPPM-09-2016-0194>.
27. Julius Murumba, Elyjoy Micheni. Big Data Analytics in Higher Education: A Review // The International Journal of Engineering and Science (IJES). 2017. Vol. 6, Issue 6. P. 14-21. <https://doi.org/10.9790/1813-0606021421>
28. Praveen Mukhia Titimus. Big Data Analytics in the Higher Education: Need of the Future. In book Advances in Communication, Cloud, and Big Data. Proceedings of 2nd National Conference on CCB 2016. Hiren Kumar Deva Sarma, Samarjeet Borah, Nitul Dutta (Eds.). Lecture Notes in Networks and Systems. Vol. 31. P. 23-28. Springer Nature Singapore Pte Ltd. 2019. https://doi.org/10.1007/978-981-10-8911-4_3
29. A. Nguyen, L. Gardner, D. Sheridan. Data Analytics in Higher Education: An Integrated View // Journal of Information Systems Education. 2020. Vol. 31, No. 1. P. 61-71. URL:
30. <https://www.researchgate.net/deref/http%3A%2F%2Fjise.org%2FVolume31%2Fn1%2FJISEv31n1p61.html>
31. Bernard Veldkamp, Kim Schildkamp, Merel Keijsers, Adrie Visscher and Ton de Jong. Big Data Analytics in Education: Big Challenges and Big Opportunities. A Transatlantic Discourse in Education Research. In book: International Perspectives on School Settings, Education Policy and Digital Strategies: A Transatlantic Discourse in Education Research, P. 266-282. Annika Wilmers, Sieglinde Jornitz (Eds.). Published by Verlag Barbara Budrich, 2021. <https://doi.org/10.2307/j.ctv1gbrzf4.19>
32. Lawal O. Yesufu. Predictive Learning Analytics in Higher Education. In book Data Analytics in Marketing, Entrepreneurship, and Innovation. Mounir Kehal, Shahira El Alfy (Eds). Boca Raton, 2021. P. 151-173. <https://doi.org/10.1201/9780429196614>
33. Katrina Sin, Loganathan Muthu. Application of Big Data in education data mining and learning analytics – a literature review // ICTACT Journal on Soft Computing. 2015. Vol. 5, No. 04. P. 1035-1049. <https://doi.org/10.21917/ijsc.2015.0145>
34. Klašnja-Milićević, M. Ivanović, Z. Budimac. Data science in education: Big data and learning analytics // Comput. Appl. Eng. Educ. 2017. P. 1–13. <https://doi.org/10.1002/cae.21844>
35. Bettina Berendt, Allison Littlejohn, Philippe Kern, Piotr Mitros, Xanthe Shacklock, Michael Blakemore. Big Data for monitoring educational systems. European Commission, B-1049 Brussels, 2017. 143 p. URL: <https://keanet.eu/wp-content/uploads/2019/09/Big-Data-for-Monitoring-Educational-Systems.pdf> (дата обращения 01.06.2021).
36. H. Luan, P. Geczy, H. Lai, J. Gobert, SJH Yang, H. Ogata, J. Balte, R. Guerra, P. Li and C-C. Tsai. Challenges and Future Directions of Big Data and Artificial Intelligence in Education // Front. Psychol. 2020. Vol. 11, article 580820. <https://doi.org/10.3389/fpsyg.2020.580820>
37. Radical Solutions and Learning Analytics Personalised Learning and Teaching Through Big Data. Lecture Notes in Educational Technology. Daniel Burgos (Ed.). Springer, 2020. 233 p. <https://doi.org/10.1007/978-981-15-4526-9>
38. Julio Ruiz-Palmero, Ernesto Colomo-Magaña, José Manuel Ríos-Ariza and Melchor Gómez-García. Big Data in Education: Perception of Training Advisors on Its Use in the Educational System // Social Sciences. 2021. Vol. 10, No. 5. article 153. URI: <http://hdl.handle.net/10486/693946> (дата обращения 01.06.2021). <https://doi.org/10.3390/socsci9040053>

39. И.Г. Захарова. Big Data и управление образовательным процессом // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. 2017. Том 3, № 1. С. 210-219. <https://doi.org/10.21684/2411-197X-2017-3-1-210-219>
40. José-Antonio Marín-Marín, Jesús López-Belmonte, Juan-Miguel Fernández-Campoy and José-María Romero-Rodríguez. Big Data in Education. A Bibliometric Review // Soc. Sci. 2019. Vol. 8, article 223. 13 p. <https://doi.org/10.3390/socsci8080223>
41. Benazir Quadir, Nian-Shing Chen, Pedro Isaias. Analyzing the educational goals, problems and techniques used in educational big data research from 2010 to 2018 // Interactive Learning Environments, 2020. <https://doi.org/10.1080/10494820.2020.1712427>
42. Катин А.В., Ершов П.С., Хохлов Ю.Е., Шапошник С.Б. Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, № 4–5. С. 259–277. https://doi.org/10.52605/16059921_2021_04_259
43. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2019. URL: https://circabc.europa.eu/sd/a/d9b1ab6e-a38f-485b-aeb5-8f7e2ce8d153/MQ_2019 ICT_ENT.pdf (дата обращения 01.06.2021).
44. Ершов П.С., Хохлов Ю.Е., Шапошник С.Б. Использование технологий работы с большими данными в российском бизнесе // Информационное общество. 2021. № 4–5. С. 300–314. https://doi.org/10.52605/16059921_2021_04_300
45. Катин А.В., Хохлов Ю.Е. Использование технологий работы с большими данными в системе государственного управления России // Информационное общество. 2021. № 4–5. С. 150–165. https://doi.org/10.52605/16059921_2021_04_150
46. Катин А.В., Хохлов Ю.Е. Использование технологий работы с большими данными в российском здравоохранении // Информационное общество. 2021. № 4–5. С. 185–199. https://doi.org/10.52605/16059921_2021_04_185
47. Малахов В.А., Хохлов Ю.Е., Шапошник С.Б. Использование технологий работы с большими данными в российской науке // Информационное общество. 2021. № 4–5. С. 200–219. https://doi.org/10.52605/16059921_2021_04_200
48. Росстат. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг» за 2020 год (размещено - 13.10.2021). URL: <https://rosstat.gov.ru/storage/mediabank/3-inform.html> (дата обращения 15.10.2021)

MONITORING THE USE OF BIG DATA TECHNOLOGIES IN THE RUSSIAN EDUCATIONAL SYSTEM

Elizarov, Alexander Mikhailovich

*Doctor of Physical and Mathematical Sciences, Professor
Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University,
Department of Software Engineering, Professor
Kazan, Russian Federation
amelizarov@gmail.com*

Hohlov, Yuri Eugenyevich

*Candidate of Physical and Mathematical Sciences, Associate Professor
Institute of the Information Society, Chairman of the Board of Directors
Plekhanov Russian University of Economics, IIS-Based Digital Economy Department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Abstract

To conduct regular monitoring the use of big data technologies in Russian educational institutions, approaches have been developed to assess the current level of readiness to use these technologies, the degree of their penetration into the education sector, to identify the existing promising areas of development, the strengths and weaknesses of this process, as well as the existing barriers. Based on the analysis of the documentary flow on the topic under consideration, carried out on Web of Science platform, it has been revealed that in international practice there are still no universal approaches and tools for monitoring the use of big data technologies in the field of education, but the issues of assessing maturity and readiness as well as analysis of existing problems and barriers to their use are relevant for the sphere of education. The foresaid determined the choice of subject areas of monitoring and the corresponding indicators.

Keywords

big data; use of big data in education; monitoring the use of big data technologies in education; the level of maturity in working with big data; barriers to using big data; BD4DE; Big Data for Digital Economy

References

1. Ukaz Prezidenta Rossijskoj Federacii ot 07.05.2018 g. № 204. O nacional'nyh celyah i strategicheskikh zadachah razvitiya Rossijskoj Federacii na period do 2024 goda. URL: <http://kremlin.ru/acts/bank/43027> (accessed on 01.06.2021).
2. Pasport nacional'nogo proekta «Obrazovanie». URL: <http://government.ru/info/35566/> (accessed on 01.06.2021); Pasport nacional'nogo proekta «Nauka». URL: <http://government.ru/info/35565/> (accessed on 01.06.2021).
3. Zasedanie Soveta po strategicheskomu razvitiyu i nacional'nym proektam ot 13.07.2020 g. URL: <http://kremlin.ru/events/president/news/63635> (accessed on 01.06.2021).
4. Analiz tekushchego sostoyaniya razvitiya cifrovoj ekonomiki v Rossii. M.: Institut razvitiya informacionnogo obshchestva, 2018. 166 s. URL: <https://iis.ru/wp-content/uploads/2020/12/DECARussia2018rus.pdf> (accessed on 01.06.2021).
5. Federal'nyj zakon ot 29.12.2012 g. № 273-FZ «Ob obrazovanii v Rossijskoj Federacii». URL: <http://kremlin.ru/acts/bank/36698> (accessed on 01.06.2021).
6. Samson Oluwaseun Fadiya, Serdar Saydam, Emeka Joshua Chukwuemeka. Big Data in Education; Future Technology Integration // The International Journal of Science & Technology. 2014. Vol. 2, Issue 8. P. 65-69. URL: https://www.researchgate.net/publication/281508011_Big_Data_in_Education_Future_Technology_Integration (accessed on 01.06.2021).
7. Athanasios S. Drigas. The Use of Big Data in Education // IJCSI International Journal of Computer Science. September 2014. Vol. 11, Issue 5, No. 1. P. 58-63. URL: <https://www.ijcsi.org/papers/IJCSI-11-5-1-58-63.pdf> (accessed on 01.06.2021).

8. Jaimin N. Undavia, Sheenal Patel, Atul Patel. Future trends and scopes of Big Data Analytics in the field of Education // International Journal of Engineering and Technology (IJET). July 2017. Vol. 9, No. 3S. P. 9-14. <https://doi.org/10.21817/ijet/2017/v9i3/170903S002>
9. Tamaro Green. Big Data in Education Technology: Improvements in International Research Design: Data Driven Testing, Analysis, and Programming Paperback. March 24, 2017. 149 p.
10. Tamaro J. Green. Big Data Analytics in Higher Education. May 2017. <https://doi.org/10.13140/RG.2.2.22565.50407>
11. C. Fischer, Z.A. Pardos, R.S. Baker et al. Mining big data in education: Affordances and challenges. Review of Research in Education. 2020. Vol. 44, No. 1. P. 130-160. <https://doi.org/10.3102/0091732X20903304>
12. M.Sh. Hajirahimova, A.S. Aliyeva. Methodologies and indicators for Big Data measurement // Sovremennye informacionnye tekhnologii i IT-obrazovanie. 2017. Tom 13, № 2. С. 81-88. <https://doi.org/10.25559/SITITO.2017.2.239>
13. Noura A. Alsheikh. Developing an integrated framework to utilize big data for higher education institutions in Saudi Arabia // International Journal of Computer Science & Information Technology (IJCSIT). February 2019, Vol. 11, No. 1. P. 31-42. <https://doi.org/10.5121/ijcsit.2019.11103>
14. Şeyhmus Aydoğdu. Educational data mining studies in Turkey: a systematic review // Turkish Online Journal of Distance Education. 2020. Vol. 21, No. 3, Article 12. Pub Date: 2020-07-01. <https://doi.org/10.17718/tojde.762046>
15. V.V. Utyomov, P.M. Gorev. Razvitie obrazovatel'nyh sistem na osnove tekhnologii Big Data // Koncept. 2018. №6. URL: <https://cyberleninka.ru/article/n/razvitie-obrazovatelnyh-sistem-na-osnove-tehnologii-big-data> (accessed on 01.06.2021).
16. B.A. Kondratenko, A.B. Kondratenko. Perspektivy primeneniya analiza bol'shih dannyh v sovremennom obrazovanii // Vestnik Baltijskogo federal'nogo universiteta im. I. Kanta. Ser.: Filologiya, pedagogika, psihologiya. 2018. № 1. S. 117-126. URL: https://journals.kantiana.ru/upload/iblock/d2a/%D0%9A%D0%BE%D0%BD%D0%B4%D1%80%D0%B0%D1%82%D0%B5%D0%BD%D0%BA%D0%BE%20%D0%91.%20%D0%90.,%20%D0%9A%D0%BE%D0%BD%D0%B4%D1%80%D0%B0%D1%82%D0%B5%D0%BD%D0%BA%D0%BE_117-126.pdf (accessed on 01.06.2021).
17. Yu.V. Malysheva. Tekhnologiya bol'shih dannyh v rossijskom obrazovanii // Nauchnye issledovaniya XXI veka. 2020. Tom 1. №3. S. 59–62.
18. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologij raboty s bol'shimi dannymi // Informacionnoe obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
19. Reestr otkrytyh dannyh Minobrnauki. <https://minobrnauki.gov.ru/opendata/> (accessed on 01.06.2021).
20. Issledovanie rynka onlajn-obucheniya 2020. URL: <http://research.edmarket.ru/> <https://minobrnauki.gov.ru/opendata/> (accessed on 01.06.2021).
21. Mezhdunarodnaya kompaniya v oblasti strategii, konsaltinga, cifrovizacii biznesa, tekhnologij i operacij Accenture. URL: <https://www.accenture.com/ru-ru> (accessed on 01.06.2021).
22. Zhurnal Computerworld Rossiya. URL: <https://www.computerworld.ru/> (accessed on 01.06.2021).
23. Associaciya bol'shih dannyh. URL: <https://rubda.ru/> (accessed on 01.06.2021).
24. Vatsala, Rutuja Jadhav, R. Sathyaraj. A Review of Big Data Analytics in Sector of Higher Education // Int. Journal of Engineering Research and Application. 2017. Vol. 7, Issue 6 (Part 2). P. 25-32. URL: <https://doi.org/10.9790/9622-0706022532> (date of the application 01.06.2021).
25. Big Data and Learning Analytics in Higher Education. Current Theory and Practice. Ben Kei Daniel (Ed.). Springer, 2017. 287 p. <https://doi.org/10.1007/978-3-319-06520-5>
26. Chad Laux, Na Li, Corey Seliger, John Springer. Impacting Big Data analytics in higher education through Six Sigma techniques // International Journal of Productivity and Performance Management. 2017. Vol. 66, Issue 5. P. 662-679, <https://doi.org/10.1108/IJPPM-09-2016-0194>.
27. Julius Murumba, Elyjoy Micheni. Big Data Analytics in Higher Education: A Review // The International Journal of Engineering and Science (IJES). 2017. Vol. 6, Issue 6. P. 14-21. <https://doi.org/10.9790/1813-0606021421>

28. Praveen Mukhia Titimus. Big Data Analytics in the Higher Education: Need of the Future. In book *Advances in Communication, Cloud, and Big Data. Proceedings of 2nd National Conference on CCB 2016*. Hiren Kumar Deva Sarma, Samarjeet Borah, Nitul Dutta (Eds.). Lecture Notes in Networks and Systems. Vol. 31. P. 23-28. Springer Nature Singapore Pte Ltd. 2019. https://doi.org/10.1007/978-981-10-8911-4_3
29. A. Nguyen, L. Gardner, D. Sheridan. Data Analytics in Higher Education: An Integrated View // *Journal of Information Systems Education*. 2020. Vol. 31, No. 1. P. 61-71. URL:
30. <https://www.researchgate.net/deref/http%3A%2F%2Fjise.org%2FVolume31%2Fn1%2FJISEv31n1p61.html>
31. Bernard Veldkamp, Kim Schildkamp, Merel Keijsers, Adrie Visscher and Ton de Jong. Big Data Analytics in Education: Big Challenges and Big Opportunities. A Transatlantic Discourse in Education Research. In book: *International Perspectives on School Settings, Education Policy and Digital Strategies: A Transatlantic Discourse in Education Research*, P. 266-282. Annika Wilmers, Sieglinde Jornitz (Eds.). Published by Verlag Barbara Budrich, 2021. <https://doi.org/10.2307/j.ctv1gbrzf4.19>
32. Lawal O. Yesufu. Predictive Learning Analytics in Higher Education. In book *Data Analytics in Marketing, Entrepreneurship, and Innovation*. Mounir Kehal, Shahira El Alfy (Eds). Boca Raton, 2021. P. 151-173. <https://doi.org/10.1201/9780429196614>
33. Katrina Sin, Loganathan Muthu. Application of Big Data in education data mining and learning analytics – a literature review // *ICTACT Journal on Soft Computing*. 2015. Vol. 5, No. 04. P. 1035-1049. <https://doi.org/10.21917/ijsc.2015.0145>
34. Klačnja-Milićević, M. Ivanović, Z. Budimac. Data science in education: Big data and learning analytics // *Comput. Appl. Eng. Educ*. 2017. P. 1–13. <https://doi.org/10.1002/cae.21844>
35. Bettina Berendt, Allison Littlejohn, Philippe Kern, Piotr Mitros, Xanthe Shacklock, Michael Blakemore. *Big Data for monitoring educational systems*. European Commission, B-1049 Brussels, 2017. 143 p.
URL: <https://keanet.eu/wp-content/uploads/2019/09/Big-Data-for-Monitoring-Educational-Systems.pdf> (accessed on 01.06.2021).
36. H. Luan, P. Geczy, H. Lai, J. Gobert, SJH Yang, H. Ogata, J. Balte, R. Guerra, P. Li and C-C. Tsai. Challenges and Future Directions of Big Data and Artificial Intelligence in Education // *Front. Psychol*. 2020. Vol. 11, article 580820. <https://doi.org/10.3389/fpsyg.2020.580820>
37. *Radical Solutions and Learning Analytics Personalised Learning and Teaching Through Big Data*. Lecture Notes in Educational Technology. Daniel Burgos (Ed.). Springer, 2020. 233 p. <https://doi.org/10.1007/978-981-15-4526-9>
38. Julio Ruiz-Palmero, Ernesto Colomo-Magaña, José Manuel Ríos-Ariza and Melchor Gómez-García. Big Data in Education: Perception of Training Advisors on Its Use in the Educational System // *Social Sciences*. 2021. Vol. 10, No. 5. article 153.
URI: <http://hdl.handle.net/10486/693946> (accessed on 01.06.2021).
<http://doi.org/10.3390/socsci9040053>
39. I.G. Zaharova. Big Data i upravlenie obrazovatel'nym processom // *Vestnik Tyumenskogo gosudarstvennogo universiteta. Gumanitarnye issledovaniya. Humanitates*. 2017. Tom 3, № 1. S. 210-219. <https://doi.org/10.21684/2411-197X-2017-3-1-210-219>
40. José-Antonio Marín-Marín, Jesús López-Belmonte, Juan-Miguel Fernández-Campoy and José-María Romero-Rodríguez. Big Data in Education. A Bibliometric Review // *Soc. Sci*. 2019. Vol. 8, article 223. 13 p. <https://doi.org/10.3390/socsci8080223>
41. Benazir Quadir, Nian-Shing Chen, Pedro Isaias. Analyzing the educational goals, problems and techniques used in educational big data research from 2010 to 2018 // *Interactive Learning Environments*, 2020. <https://doi.org/10.1080/10494820.2020.1712427>
42. A.V. Katin, P.S. Ershov, Yu.E. Hohlov, S.B. Shaposhnik. Model' BD4DE-MM zrelosti raboty s bol'shimi dannymi v organizacii // *Informacionnoe obshchestvo*. 2021, № 4–5. S. 259–277. https://doi.org/10.52605/16059921_2021_04_259
43. 42. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2019.
URL: https://circabc.europa.eu/sd/a/d9b1ab6e-a38f-485b-aeb5-8f7e2ce8d153/MQ_2019 ICT_ENT.pdf (accessed on 01.06.2021).
44. P.S. Ershov, YU.E. Hohlov, S.B. Shaposhnik. Ispol'zovanie tekhnologij raboty s bol'shimi dannymi v rossijskom biznese // *Informacionnoe obshchestvo*. 2021, № 4–5. S. 300–314. https://doi.org/10.52605/16059921_2021_04_300

45. A.V. Katin, YU.E. Hohlov. Ispol'zovanie tekhnologij raboty s bol'shimi dannymi v sisteme gosudarstvennogo upravleniya Rossii // Informacionnoe obshchestvo. 2021, № 4–5. S. 150–165. https://doi.org/10.52605/16059921_2021_04_150
46. A.V. Katin, Yu.E. Hohlov. Ispol'zovanie tekhnologij raboty s bol'shimi dannymi v rossijskom zdavoohranenii // Informacionnoe obshchestvo. 2021. 2021, № 4–5. S. 185–199. https://doi.org/10.52605/16059921_2021_04_185
47. V.A. Malahov, Yu.E. Hohlov, S.B. Shaposhnik. Ispol'zovanie tekhnologij raboty s bol'shimi dannymi v rossijskoj nauke // Informacionnoe obshchestvo. 2021. 2021, № 4–5. S. 200–219. https://doi.org/10.52605/16059921_2021_04_200
48. Rosstat. Itogi federal'nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol'zovanii cifrovyy`x tekhnologij i proizvodstve svyazanny`x s nimi tovarov i uslug» za 2020 god (razmeshheno - 13.10.2021) (accessed on 15.10.2021).

Здравоохранение в информационном обществе**МОНИТОРИНГ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ В
РОССИЙСКОЙ СИСТЕМЕ ЗДРАВООХРАНЕНИЯ****Катин Александр Владимирович**

*Институт развития информационного общества, руководитель дирекции отраслевых программ РЭУ имени Г. В. Плеханова, старший преподаватель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
alexander.katin@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Исследованы вопросы мониторинга и оценки уровня использования технологий работы с большими данными в системе здравоохранения России. Разработана концептуальная схема и система показателей для мониторинга уровня зрелости технологий работы с большими данными в сфере здравоохранения, интенсивности использования указанных технологий в медицинских организациях, а также имеющихся барьеров для их внедрения и полноценного использования. Для оценки применимости разработанной концептуальной схемы проведена пилотная реализация.

Ключевые слова

большие данные, технологии работы с большими данными, здравоохранение, использование больших данных в системе здравоохранения, мониторинг, цифровая экономика

Введение

В настоящее время сфера здравоохранения является одной из наиболее приоритетных для внедрения и использования технологий работы с большими данными, поскольку при рациональном использовании они способны повысить качество и доступность медицинских услуг, что подтверждается исследованиями ведущих международных консалтинговых компаний, в частности, McKinsey [1] и Deloitte [2].

Глобальный рынок больших данных в сфере здравоохранения в 2017 году оценивался в 16,87 млрд долларов США, а прогнозное значение на 2025 год составляет 67,82 млрд долларов США [3].

В Российской Федерации реализуется ряд инициатив, направленных на внедрение цифровых технологий в сфере здравоохранения. В соответствии с национальным проектом «Здравоохранение» [4] предусмотрена реализация федерального проекта «Создание единого цифрового контура в здравоохранении на основе единой государственной информационной системы здравоохранения (ЕГИСЗ)», основной задачей которого является создание механизмов взаимодействия медицинских организаций на основе единой государственной информационной системы в сфере здравоохранения, внедрение цифровых технологий и платформенных решений. В рамках реализации Национальной технологической инициативы утверждена [5] дорожная карта «Хелснет», которая направлена на развитие рынка персонализированных медицинских услуг и лекарственных средств, обеспечивающих рост продолжительности жизни, а также получение новых эффективных средств профилактики и лечения различных заболеваний. В России принят

© Катин А.В., Хохлов Ю.Е., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

DOI: 10.52605/16059921_2021_04_185

«Закон о телемедицине» [6], который предусматривает возможность применения телемедицинских технологий как при оказании первичной медико-санитарной помощи, так и скорой, специализированной, высокотехнологичной, паллиативной помощи пациентам и открывает широкие возможности для применения цифровых технологий в сфере здравоохранения.

Несмотря на значительный объем реализуемых проектов, сегодня технологиям работы с большими данными в сфере здравоохранения уделяется недостаточно внимания. Для того чтобы установить текущий уровень готовности к использованию технологий работы с большими данными медицинских организаций, степень их проникновения в сферу здравоохранения, выявить перспективные направления развития, сильные и слабые стороны, а также имеющиеся барьеры, необходимо регулярно проводить мониторинг уровня использования технологий работы с большими данными. Стоит отметить, что сейчас в международной практике отсутствуют универсальные подходы и инструменты мониторинга использования цифровых технологий в сфере здравоохранения, имеющиеся статистические данные ОЭСР [7] и Всемирной организации здравоохранения [8] направлены в большинстве своем на медицину, без увязки с использованием информационно-коммуникационных технологий.

Целью данной статьи является разработка концептуальной схемы и методологии проведения мониторинга и оценки уровня использования технологий работы с большими данными в системе здравоохранения Российской Федерации.

1 Предметная область мониторинга и обзор литературы

1.1 Описание предметной области мониторинга

Предметной областью мониторинга является деятельность медицинских организаций в части использования технологий работы с большими данными. В соответствии с российским законодательством [9] медицинскими организациями являются юридические лица независимо от организационно-правовой формы, осуществляющее в качестве основного (уставного) вида деятельности медицинскую деятельность на основании лицензии, выданной в порядке, установленном законодательством Российской Федерации о лицензировании отдельных видов деятельности. По данным Росстата [10], на конец 2020 года их в России насчитывалось 26 300, среди которых 5100 больничных организаций и 21200 врачебных амбулаторно-поликлинических организаций.

Помимо мониторинга уровня использования технологий работы с большими данными в деятельности медицинских организаций, видится целесообразным включить в предметную область мониторинга аспекты, позволяющие исследовать степень готовности российских медицинских организаций к полноценному внедрению и использованию технологий больших данных, а также имеющихся препятствий для их применения.

1.2 Обзор научных публикаций

В целях формирования концептуальной схемы мониторинга и оценки уровня использования технологий работы с большими данными в российской системе здравоохранения авторами проведен отбор и анализ релевантных научных публикаций, индексируемых платформой Web of Science (WoS). Для этого из «ядерной» коллекции WoS Core Collection был выделен массив публикаций по большим данным, определенный по следующему поисковому предписанию:

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Подробное описание процесса разработки поискового запроса приведено в статье [11].

За временной период с 2016 по 2020 год посредством поискового образа по полю «Тема» (поиск по названию, аннотации, автору и ключевым словам) было найдено 108 073 публикации.

На следующем этапе из этого найденного массива осуществлялся отбор публикаций, имеющих отношение к системе здравоохранения, путем задания специализированного поискового образа, содержащего ключевые слова и словосочетания, относящиеся к сфере здравоохранения:

healthcare OR health OR telemedicine OR treatment OR “healthcare organization” OR “medical organization” OR “medical care” OR “digital healthcare” OR “healthcare digitalization” OR “healthcare industry” OR “preventive care” OR “patient record analysis” OR hospital OR clinic OR treating OR “digital transformation in healthcare” OR “medical innovation” OR “healthcare management” OR “medical diagnostic” OR “precision medicine” OR “electronic health” OR e-health OR e-healthcare OR “medical records” OR “medical data” OR “medical service” OR “instantaneous healthcare” OR “in-depth healthcare”

Приведенные ключевые слова с одной стороны позволяют получить максимальное покрытие исследуемой области, а с другой – не дают значительного «шума», то есть публикаций, слабо связанных с темой исследования (подбор ключевых слов осуществлялся методом экспертных оценок с последующей верификацией).

Общий и специализированный поисковые запросы были объединены логическим оператором “AND” и в результате, в рамках поиска по полю «Тема», было получено 16975 публикации за временной период с 2016 по 2020 год.

Исходя из того, что полученный массив оказался достаточно обширным и практически не поддающимся анализу, было принято решение сосредоточиться на обзорных статьях, которых оказалось 1647.

На следующем этапе были изучены наименования и аннотации полученных работ, в результате чего были отобраны 25 наиболее релевантных теме исследования публикаций [12-36], анализ полного текста которых позволил бы сформировать концептуальную схему мониторинга и оценки уровня использования технологий работы с большими данными в системе здравоохранения. Основным критерием для отбора было наличие в публикации вопросов, относящихся к использованию технологий работы с большими данными в медицинских организациях при оказании медицинских услуг или для управления организацией. Еще одним критерием, позволяющим говорить о необходимости изучения полного текста статьи, являлось исследование в статье факторов, влияющих на использование технологий работы с большими данными в системе здравоохранения, таких как: стратегическое планирование, регулирование, человеческих капитал, цифровая инфраструктура, доверие и безопасность.

Большая часть работ [12-23] посвящена обзорам применения технологий работы с большими данными в деятельности организаций сферы здравоохранения как для оказания медицинских услуг, так и для управления медицинскими организациями или системой здравоохранения в целом. Например, в работе [12] приводится обзор приложений, используемых для интеллектуального анализа данных и прогнозной аналитики при оказании медицинских услуг, а в работе [16] приводится обзор инструментов, основанных на работе с большими данными и направленных на управление процессами изобретения и оборота лекарственных средств.

В отобранном перечне публикаций имеются статьи, посвященные обзору вопросов, связанных с готовностью медицинских организаций внедрять и использовать технологии работы с большими данными [24-31]. В частности, в работе [25] исследованы вопросы воздействия технологий работы с большими данными на деятельность медицинских организаций, в статье [24] рассмотрена необходимость внесения изменений в методическое обеспечение проведения медицинских процедур в связи с использованием технологий работы с большими данными.

Также важным направлением исследований являются препятствия, имеющиеся в сфере здравоохранения и отдельных медицинских организациях для использования технологий работы с большими данными [32-36].

Анализ полученных в результате проведенного отбора релевантных научных публикаций показал, что в текущий момент времени в международной практике отсутствуют универсальные, общепринятые подходы к мониторингу уровня использования технологий работы с большими данными в сфере здравоохранения. Вместе с тем в большинстве обзорных работ, упомянутых выше, в той или иной форме говорится об использовании отдельных инструментов работы с большими данными в медицинских организациях, зрелости и готовности к работе с большими данными в сфере здравоохранения, а также об имеющихся препятствиях для их использования. Исходя из этого и была сформирована концептуальная схема мониторинга и оценки уровня использования технологий работы с большими данными в системе здравоохранения России.

2 Концептуальная схема и показатели мониторинга

Концептуальная схема мониторинга использования технологий для работы с большими данными в российском здравоохранении представлена на рисунке 1.

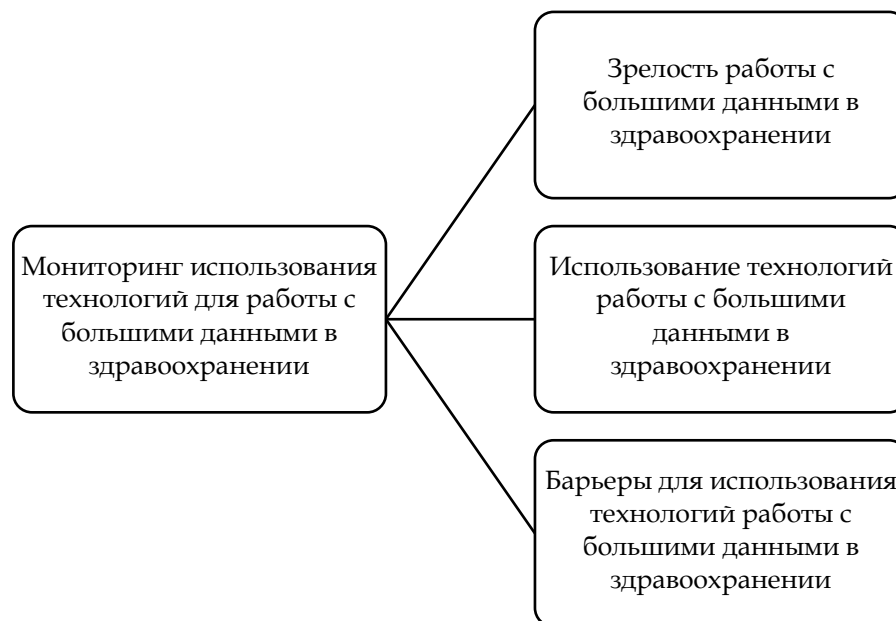


Рисунок 1. Концептуальная схема мониторинга использования технологий для работы с большими данными в системе здравоохранения России

2.1 Зрелость работы с большими данными в здравоохранении

Оценка уровня зрелости работы с большими данными позволит установить степень готовности медицинских организаций к работе с большими данными, выявить сильные и слабые стороны. Наличие интегральной оценки для отрасли позволит более рационально и обоснованно выстраивать государственную политику в сфере развития технологий работы с большими данными и внедрения их в сферу здравоохранения. Для медицинских организаций наличие инструмента оценки их зрелости позволит повысить эффективность своей деятельности в части внедрения и использования цифровых технологий, включая технологии работы с большими данными. Для оценки зрелости работы с большими данными будет использован ряд показателей, характеризующих интегральный уровень зрелости российской сферы здравоохранения, а также показатели, позволяющие оценить степень зрелости по отдельно взятым аспектам.

В основу используемого в настоящей статье подхода легла модель BD4DE-MM зрелости работы с большими данными в организации» [37].

- (Ц33-01) Доля медицинских организаций, имеющих продвинутый и выше уровень зрелости работы с большими данными

Данный показатель позволяет дать интегральную оценку уровня зрелости работы с большими данными в российской системе здравоохранения. Низкое значение показателя будет говорить о недостаточной готовности к внедрению и использованию технологий работы с большими данными большинства медицинских организаций в России. Показатель может быть полезен лицам, принимающим решения по цифровой трансформации отечественной сферы здравоохранения, а ежегодный рост значения показателя может свидетельствовать о правильности проводимой политики.

- (Ц33-01-01) Доля медицинских организаций, имеющих продвинутый и выше уровень зрелости стратегии и регулирования работы с большими данными

Показатель характеризует наличие в медицинской организации стратегии использования технологий работы с большими данными, а также качество ее проработки (согласованность с общей стратегией развития организации, наличие плана реализации стратегии, обеспеченность ресурсами, учет актуальных тенденций). Также показатель позволяет оценить уровень

вовлеченности организации в нормативное правовое и техническое регулирование использования технологий работы с большими данными в сфере здравоохранения.

- *(ЦЗЗ-01-02) Доля медицинских организаций, имеющих продвинутый и выше уровень кадрового обеспечения работы с большими данными*

Показатель позволяет оценить такую важную составляющую процесса внедрения и использования технологий работы с большими данными, как наличие кадров, уровень их знаний и компетенций. Соответствующими навыками должны обладать как врачи, так и административные руководители медицинской организации. Кроме того, для получения максимальной оценки, в медицинской организации должны иметься лидеры, которые способны инициировать и возглавить процесс внедрения технологий работы с большими данными во все деловые процессы организации

- *(ЦЗЗ-01-03) Доля медицинских организаций, имеющих продвинутый и выше уровень обеспеченности качественными большими данными*

Показатель направлен на мониторинг и оценку уровня обеспеченности всех деловых процессов медицинской организации качественными большими данными, подлежащими совместному использованию.

- *(ЦЗЗ-01-04) Доля медицинских организаций, имеющих продвинутый и выше уровень использования аналитики и инструментов для работы с большими данными*

Показатель характеризует уровень использования медицинской организацией аналитики и инструментов для работы с большими данными. Использование технологий работы с большими данными позволяет значительно повысить эффективность системы здравоохранения по множеству направлений, среди которых прогнозирование течения болезни, планирование лечения, сокращение повторных госпитализаций, оптимизация затрат и повышение качества медицинского обслуживания.

- *(ЦЗЗ-01-05) Доля медицинских организаций, имеющих продвинутый и выше уровень зрелости цифровой инфраструктуры и безопасности для работы с большими данными*

Наличие в медицинской организации современной цифровой инфраструктуры является базовым условием для полноценного использования технологий работы с большими данными. Элементы инфраструктуры могут принадлежать медицинской организации, либо быть арендованными по модели IaaS (Infrastructure-as-a-Service, «инфраструктура как сервис»), однако они должны отвечать требованиям, зачастую достаточно высоким, которые к ней предъявляются технологиями работы с большими данными.

Вопросы безопасности при использовании больших данных в сфере здравоохранения являются одними из важнейших, поскольку медицинские организации оперируют персональными данными пользователей, включая медицинские данные, составляющие медицинскую тайну.

Кроме того, ненадежное функционирование инфраструктуры, а также ненадлежащая защита данных могут привести к блокировке функционирования медицинской организации, что несет за собой серьезные риски для здоровья и благополучия пациентов.

- *(ЦЗЗ-01-06) Доля медицинских организаций, имеющих продвинутый и выше уровень зрелости организационных механизмов работы с большими данными*

Показатель характеризует уровень развитости организационных механизмов работы с большими данными, что выражается в наличие ответственного подразделения и руководителя, а также документов, регламентирующих использование технологий работы с большими данными в медицинской организации.

Показатель ЦЗЗ-01, позволяющий оценить уровень зрелости работы с большими данными всей системы здравоохранения России, равно как и показатели ЦЗЗ-01-01 – ЦЗЗ-01-06, оценивающие отдельные области зрелости, является универсальным и может быть использован (с учетом имеющейся специфики) для других сфер деятельности, что дает возможность обоснованного сопоставления. Следует отметить, что применительно к системе здравоохранения, основное внимание должно уделяться процессам оказания медицинских услуг, а также процессам управления медицинской организацией.

2.2 Использование технологий работы с большими данными в здравоохранении

Мониторинг уровня использования технологий работы с большими данными в сфере здравоохранения позволит оценить уровень их проникновения в деятельность медицинских организаций, что даст возможность выявить имеющиеся приоритеты, сравнить достигнутый уровень с другими сферами деятельности. Наличие такой оценки, регулярно корректируемой в рамках мониторинга, поможет компаниям – поставщикам технологий работы с большими данными более эффективно выстраивать свою деятельность в части формирования предложения для сферы здравоохранения.

- *(ИБДЗ-01) Доля медицинских организаций, использующих технологии работы с большими данными*

Показатель позволяет оценить уровень использования технологий работы с большими данными в сфере здравоохранения в разрезе стадий жизненного цикла данных, а именно: при генерации больших данных, их сборе, обработке, хранении, анализе и визуализации.

- *(ИБДЗ-02) Доля медицинских организаций, осуществляющих работу с большими данными собственными силами или силами внешних компаний*

Показатель позволяет оценить с одной стороны имеющийся в медицинских организациях уровень компетенций и возможностей, а с другой, оценить спрос на услуги в сфере работы с большими данными в системе здравоохранения.

- *(ИБДЗ-03) Доля медицинских организаций, предоставляющих (продающих), а также получающих (покупающих) доступ к большим данным*

Сфера здравоохранения генерирует значительный объем разного формата данных, из множества источников (данные о пациентах, результаты медицинских исследований, данные носимых пациентами устройств и т. д.). Все эти данные могут быть полезны как компаниям, которые осуществляют разработки в сфере работы с большими данными, так и самим медицинским организациям, которые используют их для повышения качества оказываемых услуг. Показатель позволяет оценить, насколько эти данные востребованы, а также какова активность медицинских организаций на рынке больших данных.

- *(ИБДЗ-04) Доля затрат медицинских организаций на работу с большими данными в общих затратах на цифровые технологии*

В России выделяются значительные бюджетные средства на цифровую трансформацию сферы здравоохранения. В рамках реализации Единой государственной информационной системы здравоохранения (далее – ЕГИСЗ), на реализацию одних только региональных проектов в 85 субъектах Российской Федерации выделено более 81 млрд рублей из федерального бюджета, а консолидированные бюджеты регионов на создание и внедрение региональных сегментов ЕГИСЗ составят более 88 млрд рублей. Однако стоит отметить, что выделить из общего бюджета на внедрение цифровых технологий в медицинских организациях затраты на технологии работы с большими данными не представляется возможным. Поэтому, в рамках мониторинга, предлагается ввести показатель «Доля затрат медицинских организаций на работу с большими данными в общих затратах на цифровые технологии», который позволит хотя бы приблизительно оценить рынок технологий работы с большими данными в российской сфере здравоохранения в денежном выражении.

- *(ИБДЗ-05) Доля медицинских организаций, применяющих аналитику больших данных (в разрезе по целям использования)*

Важным аспектом использования технологий работы с большими данными является применение аналитики. В системе здравоохранения анализ больших данных может использоваться в целях повышения качества оказываемых медицинских услуг, либо для повышения эффективности управления медицинским учреждением. В рамках настоящего мониторинга видится необходимым установить долю медицинских организаций, применяющих аналитику больших данных, а также цели ее применения, среди которых предлагается выделить следующие:

- оказание медицинских услуг;
- управление медицинской организацией;
- маркетинг и привлечение пациентов;
- логистика;
- управление кадрами
- обеспечение информационной безопасности.

2.3 Барьеры для использования технологий работы с большими данными в здравоохранении

Мониторинг и анализ барьеров для использования технологий работы с большими данными в сфере здравоохранения является ключевым блоком исследования, поскольку только при условии минимизации или устранения препятствий для полноценного использования технологий можно говорить о потенциальных эффектах от их использования как для граждан и отдельных медицинских организаций, так и для сферы здравоохранения и страны в целом. Наличие обоснованной оценки причин неиспользования технологий работы с большими данными будет полезно для всех сторон, ответственных за развитие российской медицины, а также для бизнеса, который сможет рационально реагировать на имеющиеся на рынке проблемы и возможности. Данная область мониторинга является значимой, поскольку для полноценного развития технологий работы с большими данными все барьеры для их внедрения должны быть нивелированы или устранены.

- (ББДЗ-01) Доля медицинских организаций, считающих, что затраты на технологии работы с большими данными слишком высоки

Одним из основных барьеров является ограниченность финансовых ресурсов, выделяемых на здравоохранение, и относительно небольшая величина рынка платных медицинских услуг в России по сравнению со средствами бюджета, выделяемыми на данную сферу. Необходимо оценить долю медицинских организаций, считающих, что затраты на технологии работы с большими данными слишком высоки по сравнению с выгодами, которые потенциально могут быть получены.

- (ББДЗ-02) Доля медицинских организаций, считающих, что у них недостаточно человеческих ресурсов, знаний и навыков

Серьезным препятствием для развития цифрового здравоохранения является нехватка квалифицированных кадров, как среди медицинского, так и среди вспомогательного и административного персонала. Также для российской системы здравоохранения характерна проблема неконкурентного уровня заработной платы в медицинских организациях по сравнению с частным сектором, что затрудняет процесс привлечения высококвалифицированных специалистов в сфере больших данных. В рамках мониторинга барьеров для использования технологий работы с большими данными, необходимо оценить долю медицинских организаций, считающих, что у них недостаточно человеческих ресурсов, знаний и навыков.

- (ББДЗ-03) Доля медицинских организаций, считающих, что их цифровая инфраструктура недостаточно развита для использования технологий работы с большими данными

Критическим барьером для полноценного внедрения и использования технологий работы с большими данными является отсутствие в медицинских организациях современной цифровой инфраструктуры. В России реализуется ряд мер по наращиванию потенциала медицинских организаций в части использования цифровых технологий. В частности, в рамках реализации федерального проекта «Информационная инфраструктура» [38] реализуется мероприятие, целью которого является обеспечение к 2021 году доступа к интернету 100% медицинских организаций. В процессе мониторинга необходимо оценить долю организаций, считающих, что их цифровая инфраструктура недостаточно развита для использования технологий работы с большими данными, которые требуют высокой скорости и надежности интернет-соединения и больших вычислительных мощностей.

- (ББДЗ-04) Доля медицинских организаций, считающих, что технологии работы с большими данными не являются приоритетом для их деятельности

Применение технологий работы с большими данными является относительно новой деятельностью для российских медицинских организаций. Несмотря на их значительный потенциал, сегодня они не являются приоритетными. Для мониторинга уровня готовности медицинских организаций к использованию технологий работы с большими данными, необходимо проводить оценку доли организаций, считающих, что они не являются приоритетом для их деятельности.

- (ББДЗ-05) Доля медицинских организаций, оценивающих недостаточными источники больших данных для решения текущих задач

Полноценное функционирование технологий работы с большими данными невозможно без достаточного объема данных, а также наличия значительного количества диверсифицированных

источников их возникновения. Данный показатель позволяет оценить долю медицинских организаций, испытывающих дефицит в источниках больших данных.

- (ББДЗ-06) Доля медицинских организаций, считающих, что качество имеющихся больших массивов данных (как в организации, так и за ее пределами) не позволяет решать текущие задачи

Еще одним обязательным фактором, влияющим на возможность использования технологий работы с большими данными, является требуемый уровень качества данных. Под качеством понимается соответствие заданным характеристикам, позволяющее обрабатывать и анализировать большие данные, извлекая все возможные преимущества. Недостаточный уровень качества больших данных может привести к серьезным финансовым и временным потерям. Показатель позволяет оценить долю медицинских организаций, отказывающихся от использования технологий работы с большими данными из-за низкого качества данных.

- (ББДЗ-07) Доля медицинских организаций, полагающих, что технологии работы с большими данными бесполезны

В силу того, что технологии работы с большими данными являются достаточно инновационными, в особенности для сферы здравоохранения, заинтересованным сторонам, включая государство и бизнес, необходимо осуществлять мероприятия по информированию лиц, принимающих решения в медицинских организациях о возможностях технологий работы с большими данными. Показатель ББДЗ-06 позволяет оценить необходимые масштабы такой работы.

- (ББДЗ-08) Доля медицинских организаций, отмечающих трудности соблюдения законодательства о персональных данных при работе с большими данными

Сфера здравоохранения хранит и обрабатывает значительные объемы персональных данных, включающих медицинскую тайну. В России данная деятельность достаточно серьезно регулируется. Показатель позволяет оценить долю организаций, не внедряющих технологии работы с большими данными из опасений нарушения российского законодательства.

- (ББДЗ-09) Доля медицинских организаций, считающих использование аналитики больших данных неэтичным

Сфера здравоохранения традиционно требует соответствия высоким этическим и моральным стандартам, поскольку напрямую влияет на качество жизни граждан. Широкое внедрение цифровых технологий в систему здравоохранения у многих людей вызывает опасения. Даже гипотетическая возможность компрометации данных пациента или постановки неправильного диагноза может являться серьезным барьером для использования технологий работы с большими данными в процессы оказания медицинских услуг. Показатель позволяет оценить долю медицинских организаций, считающих использование аналитики данных в своей деятельности неэтичным.

3 Результаты пилотной реализации мониторинга

По результатам первой стадии пилотной реализации предложенной концептуальной схемы мониторинга уровня использования технологий работы с большими данными в системе здравоохранения была разработана анкета для обследования медицинских организаций.

На второй стадии пилотной реализации системы мониторинга проводилась оценка разработанной анкеты экспертом в данной предметной области (специалистом по применению цифровых технологий в здравоохранении) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций).

Экспертом по применению цифровых технологий в здравоохранении были сделаны предложения по уточнению формулировок для учета специфики данной предметной области. Также эксперт обратил внимание на то, что часть вопросов и соответствующих им показателей значительно опережают текущий уровень использования технологий работы с большими данными медицинскими организациями. Экспертом-социологом был сформулирован ряд предложений по уточнению и разъяснению для респондентов отдельных положений анкеты.

На последней стадии пилотной реализации полученные замечания и предложения экспертов были учтены в финальной версии анкеты для обследования медицинских организаций.

На основании официальных статистических данных за 2020 г. [40] были рассчитаны значения для показателей ИБДЗ-01 и, частично, для ИБДЗ-05 (рисунок 2).

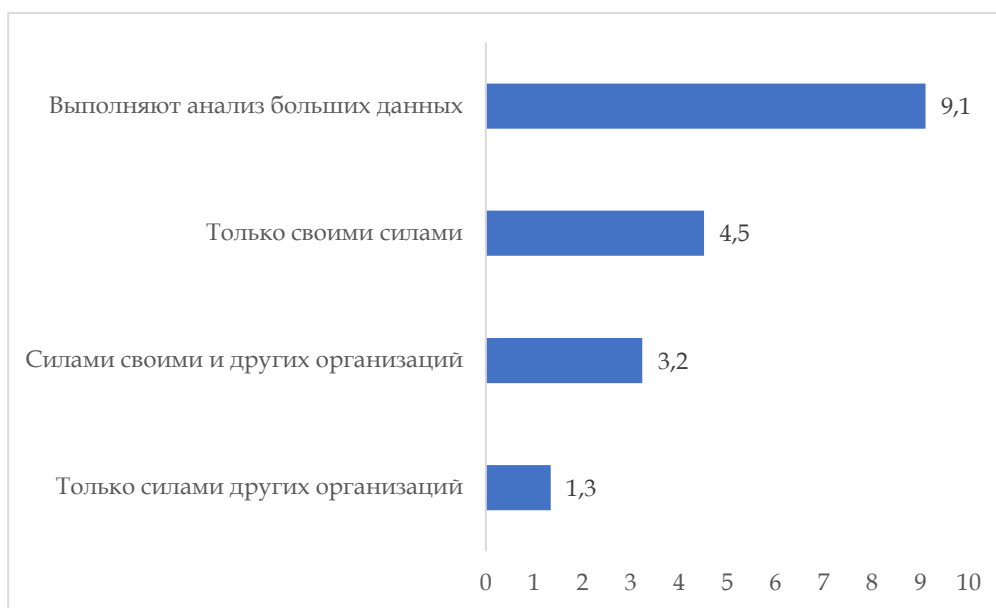


Рисунок 2. Доля медицинских организаций, выполняющих анализ больших данных, 2020, %
Источник: [40], расчеты авторов

Доля медицинских организаций, использовавших технологии сбора, обработки и анализа больших данных в 2020 году, составила 27,2%. Выполняют анализ больших данных существенно меньше медицинских организаций – 9,1 %, причем 4,5% осуществляют данную деятельность только собственными силами.

Заключение

В данной работе описана методология, концептуальная схема и показатели мониторинга и оценки уровня использования технологий работы с большими данными в системе здравоохранения России.

Применимость предлагаемой методологии оценена в рамках пилотной реализации и считается обоснованной с учетом рекомендованных доработок. Следующим этапом апробации предложенного подхода должно стать масштабное исследование, включающее в себя опрос медицинских организаций с использованием разработанной анкеты. Данные для расчета отдельных показателей по использованию технологий работы с большими данными в системе здравоохранения России будут доступны осенью 2021 года, когда Росстат опубликует результаты федерального статистического наблюдения, проведенного по обновленной форме №3-информ [39].

Дальнейшее развитие настоящего исследования может идти в направлении уточнения концептуальной схемы мониторинга, а также пересмотре перечня показателей, который может потребоваться после сбора всех необходимых данных для проведения оценки, а также получения обратной связи от опрашиваемых организаций и профессионального сообщества.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по

договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Литература

1. McKinsey & Company. The role of big data in medicine. November 2015 URL: <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine> (дата обращения: 01.06.2021).
2. Deloitte. The future of health analytics: unlocking clinical and business value. 2015. URL: <https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/future-of-health-analytics.html> (дата обращения: 01.06.2021).
3. Allied Market Research. Big Data Analytics in Healthcare Market. December 2018. URL: <https://www.alliedmarketresearch.com/big-data-analytics-in-healthcare-market> (дата обращения: 01.06.2021).
4. Паспорт национального проекта «Здравоохранение». Утвержден президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам 24 декабря 2018 года. URL: <http://government.ru/info/35561/> (дата обращения: 01.06.2021).
5. Дорожная карта «Хелснет». Одобрена Межведомственной рабочей группой по разработке и реализации Национальной технологической инициативы при Правительственной комиссии по модернизации экономики и инновационному развитию России 21.01.2021, Протокол №1. URL: <https://nti2035.ru/markets/healthnet> (дата обращения: 01.06.2021)
6. Федеральный закон от 29 июля 2017 г. N 242-ФЗ "О внесении изменений в отдельные законодательные акты Российской Федерации по вопросам применения информационных технологий в сфере охраны здоровья".
7. OECD Data. Health. URL: <https://data.oecd.org/health.htm> (дата обращения: 01.06.2021)
8. World Health Organization. Data collections. URL: <https://www.who.int/data/collections> (дата обращения: 01.06.2021).
9. Федеральный закон "Об основах охраны здоровья граждан в Российской Федерации" от 21.11.2011 N 323-ФЗ (последняя редакция).
10. Росстат. Статистические данные в сфере здравоохранения. URL: <https://www.gks.ru/folder/13721> (дата обращения: 01.06.2021).
11. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
12. Malik, M.M., Abdallah, S. & Ala'raj, M. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Ann Oper Res* 270, 287–312 (2018). <https://doi.org/10.1007/s10479-016-2393-z>
13. Salazar-Reyna, R., Gonzalez-Aleu, F., Granda-Gutierrez, E.M.A., Diaz-Ramirez, J., Garza-Reyes, J.A. and Kumar, A. (2020), "A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems", *Management Decision*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/MD-01-2020-0035>
14. G. Li, Y. Liu and H. Cai, "Research on Application of Big Data in Medical Industry," 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE), 2018, pp. 763-765, <https://doi.org/10.1109/ICSCSE.2018.00164>.
15. X. Jiao, B. Jing and B. Jiao, "Research of Prognostic and Health Management for Avionics System Based on Massive Data Mining," 2018 Prognostics and System Health Management Conference (PHM-Chongqing), 2018, pp. 350-357, <https://doi.org/10.1109/PHM-Chongqing.2018.00066>.
16. Hao Zhu. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual Review of Pharmacology and Toxicology*. Vol. 60:573-589 (Volume publication date January 2020) <https://doi.org/10.1146/annurev-pharmtox-010919-023324>
17. Chan, C.-L.; Chang, C.-C. Big Data, Decision Models, and Public Health. *Int. J. Environ. Res. Public Health* 2020, 17, 6723. <https://doi.org/10.3390/ijerph17186723>

18. Gunasekar Thangarasu, Kayalvizhi Subramanian. Big Data Analytics for Improved Care Delivery in the Healthcare Industry. *International Journal of Online and Biomedical Engineering (iJOE)* 15(10):40 <https://doi.org/10.3991/ijoe.v15i10.10875>
19. Benke, K.; Benke, G. Artificial Intelligence and Big Data in Public Health. *Int. J. Environ. Res. Public Health* 2018, 15, 2796. <https://doi.org/10.3390/ijerph15122796>
20. Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in Clinical Medicine. *Chin Med J* 2016; 129:731-8.
21. Kaplan, Bonnie, How Should Health Data Be Used? Privacy, Secondary Use, and Big Data Sales (August 1, 2014). Yale University Institute for Social and Policy Studies Working Paper No. 14-025 *Cambridge Quarterly of Healthcare Ethics* 25(2): 312-329, 2016, Available at SSRN: <https://ssrn.com/abstract=2510013> or <http://dx.doi.org/10.2139/ssrn.2510013>
22. Z. A. Mdaghri, M. El Yadari, A. Benyoussef and A. El Kenz, "Study and analysis of data mining for healthcare," 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016, pp. 77-82, <https://doi.org/10.1109/CIST.2016.7804994>.
23. C. Pasupathi and V. Kalavakonda, "Evidence Based health care system using Big Data for disease diagnosis," 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2016, pp. 743-747, <https://doi.org/10.1109/AEEICB.2016.7538393>.
24. U. Akhtar, J. W. Lee, H. S. Muhammad Bilal, T. Ali, W. A. Khan and S. Lee, "The Impact of Big Data In Healthcare Analytics," 2020 International Conference on Information Networking (ICOIN), 2020, pp. 61-63, <https://doi.org/10.1109/ICOIN48656.2020.9016588>.
25. Benabderrahmane S. (2017) What Can the Big Data Eco-System and Data Analytics Do for E-Health? A Smooth Review Study. In: Rojas I., Ortuño F. (eds) *Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science*, vol 10208. Springer, Cham. https://doi.org/10.1007/978-3-319-56148-6_56
26. Caliebe, A., Leverkus, F., Antes, G. et al. Does big data require a methodological change in medical research?. *BMC Med Res Methodol* 19, 125 (2019). <https://doi.org/10.1186/s12874-019-0774-0>
27. Dolley S (2018) Big Data's Role in Precision Public Health. *Front. Public Health* 6:68. <https://doi.org/10.3389/fpubh.2018.00068>
28. L. Elezabeth, V. P. Mishra and J. Dsouza, "The Role of Big Data Mining in Healthcare Applications," 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2018, pp. 256-260, <https://doi.org/10.1109/ICRITO.2018.8748434>
29. P. Halder and I. Pan, "Role of Big Data Analysis in Healthcare Sector: A survey," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 221-225, <https://doi.org/10.1109/ICRCICN.2018.8718684>
30. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform.* 2018;15(3):20170030. Published 2018 May 10. <https://doi.org/10.1515/jib-2017-0030>
31. Wang, Y., et al., Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, *Technol. Forecast. Soc. Change* (2016), <http://dx.doi.org/10.1016/j.techfore.2015.12.019>
32. Adnan K., Akbar R., Khor S.W., Ali A.B.A. (2020) Role and Challenges of Unstructured Big Data in Healthcare. In: Sharma N., Chakrabarti A., Balas V. (eds) *Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing*, vol 1042. Springer, Singapore. https://doi.org/10.1007/978-981-32-9949-8_22
33. Favaretto, M., De Clercq, E. & Elger, B.S. Big Data and discrimination: perils, promises and solutions. A systematic review. *J Big Data* 6, 12 (2019). <https://doi.org/10.1186/s40537-019-0177-4>
34. Hariri, R.H., Fredericks, E.M. & Bowers, K.M. Uncertainty in big data analytics: survey, opportunities, and challenges. *J Big Data* 6, 44 (2019). <https://doi.org/10.1186/s40537-019-0206-3>
35. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract.* 2017;36(1):3-11. <https://doi.org/10.23876/j.krcp.2017.36.1.3>
36. Sokolova M., Matwin S. (2016) Personal Privacy Protection in Time of Big Data. In: Matwin S., Mielniczuk J. (eds) *Challenges in Computational Statistics and Data Mining. Studies in Computational Intelligence*, vol 605. Springer, Cham. https://doi.org/10.1007/978-3-319-18781-5_18

37. Катин А.В., Ершов П.С., Хохлов Ю.Е., Шапошник С.Б. Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, № 4–5. С. 259–277. https://doi.org/10.52605/16059921_2021_04_259
38. Паспорт федерального проекта «Информационная инфраструктура». Утвержден президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 28 мая 2019 г. № 9)
39. Форма № 3-информ. Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг. Утверждена Приказом Росстата «Об утверждении формы» от 30.07.2020 № 424.
40. Росстат. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг» за 2020 год (размещено - 13.10.2021). URL: <https://rosstat.gov.ru/storage/mediabank/3-inform.html> (дата обращения 15.10.2021).

MONITORING THE USE OF BIG DATA IN THE RUSSIAN HEALTHCARE SYSTEM

Katin, Alexander Vladimirovich

*Institute of the Information Society, head of Directorate of sectoral programs
Plekhanov Russian University of Economics, IIS-based Digital economy department, senior lecturer
Moscow, Russian Federation
alexander.katin@iis.ru*

Hohlov, Yuri Eugenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-Based Digital Economy Department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Abstract

The issues of monitoring and assessing the level of big data technologies use in the healthcare system of Russia have been investigated. A conceptual scheme and a system of indicators have been developed to monitor the level of maturity of big data technologies in the healthcare sector, the intensity of these technologies use in medical organizations, as well as the existing barriers to their implementation and full usage. To assess the applicability of the developed conceptual scheme, a pilot implementation was carried out.

Keywords

big data, big data technologies, healthcare, the use of big data in the healthcare system, monitoring, digital economy

References

1. McKinsey & Company. The role of big data in medicine. November 2015 [Electronic resource] URL: <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine> (accessed on 01.06.2021)
2. Deloitte. The future of health analytics: unlocking clinical and business value. 2015. [Electronic resource] URL: <https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/future-of-health-analytics.html> (accessed on 01.06.2021)
3. Allied Market Research. Big Data Analytics in Healthcare Market. December 2018. [Electronic resource] URL: <https://www.alliedmarketresearch.com/big-data-analytics-in-healthcare-market> (accessed on 01.06.2021)
4. Passport natsionalnogo proekta «Zdravookhranenie». Utverzhden prezidiumom Soveta pri Prezidente Rossiiskoi Federatsii po strategicheskomu razvitiyu i natsionalnym proektam 24 dekabrya 2018 goda. [Electronic resource] URL: <http://government.ru/info/35561/> (accessed on 01.06.2021)
5. Dorozhnaya karta «Khelsnet». Odobrena Mezhhvedomstvennoi rabochei gruppoi po razrabotke i realizatsii Natsionalnoi tekhnologicheskoi initsiativy pri Pravitelstvennoi komissii po modernizatsii ekonomiki i innovatsionnomu razvitiyu Rossii 21.01.2021, Protokol №1. [Electronic resource] URL: https://nti2035.ru/markets/healthnet_ (accessed on 01.06.2021)
6. Federalnyi zakon ot 29 iyulya 2017 g. № 242-FZ "O vnesenii izmenenii v otдельnye zakonodatelnye akty Rossiiskoi Federatsii po voprosam primeneniya informatsionnykh tekhnologii v sfere okhrany zdorovya"
7. OECD Data. Health. [Electronic resource] URL: <https://data.oecd.org/health.htm> (accessed on 01.06.2021)
8. World Health Organization. Data collections. [Electronic resource] URL: <https://www.who.int/data/collections> (accessed on 01.06.2021)
9. Federalnyi zakon "Ob osnovakh okhrany zdorovya grazhdan v Rossiiskoi Federatsii" ot 21.11.2011 № 323-FZ (poslednyaya redaktsiya)
10. Rosstat. Statisticheskie dannye v sfere zdravookhraneniya. [Electronic resource] URL: <https://www.gks.ru/folder/13721> (accessed on 01.06.2021)

11. T.V. Ershova, Yu.E. Khokhlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispolzovaniya tekhnologii raboty s bolshimi dannymi // Informatsionnoe obshchestvo. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
12. Malik, M.M., Abdallah, S. & Ala'raj, M. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Ann Oper Res* 270, 287–312 (2018). <https://doi.org/10.1007/s10479-016-2393-z>
13. Salazar-Reyna, R., Gonzalez-Aleu, F., Granda-Gutierrez, E.M.A., Diaz-Ramirez, J., Garza-Reyes, J.A. and Kumar, A. (2020), "A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems", *Management Decision*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/MD-01-2020-0035>
14. G. Li, Y. Liu and H. Cai, "Research on Application of Big Data in Medical Industry," 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE), 2018, pp. 763-765, <https://doi.org/10.1109/ICSCSE.2018.00164>
15. X. Jiao, B. Jing and B. Jiao, "Research of Prognostic and Health Management for Avionics System Based on Massive Data Mining," 2018 Prognostics and System Health Management Conference (PHM-Chongqing), 2018, pp. 350-357, <https://doi.org/10.1109/PHM-Chongqing.2018.00066>.
16. Hao Zhu. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual Review of Pharmacology and Toxicology*. Vol. 60:573-589 (Volume publication date January 2020) <https://doi.org/10.1146/annurev-pharmtox-010919-023324>
17. Chan, C.-L.; Chang, C.-C. Big Data, Decision Models, and Public Health. *Int. J. Environ. Res. Public Health* 2020, 17, 6723. <https://doi.org/10.3390/ijerph17186723>
18. Gunasekar Thangarasu, Kayalvizhi Subramanian. Big Data Analytics for Improved Care Delivery in the Healthcare Industry. *International Journal of Online and Biomedical Engineering (ijOE)* 15(10):40 <https://doi.org/10.3991/ijoe.v15i10.10875>
19. Benke, K.; Benke, G. Artificial Intelligence and Big Data in Public Health. *Int. J. Environ. Res. Public Health* 2018, 15, 2796. <https://doi.org/10.3390/ijerph15122796>
20. Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in Clinical Medicine. *Chin Med J* 2016; 129:731-8
21. Kaplan, Bonnie, How Should Health Data Be Used? Privacy, Secondary Use, and Big Data Sales (August 1, 2014). Yale University Institute for Social and Policy Studies Working Paper No. 14-025 *Cambridge Quarterly of Healthcare Ethics* 25(2): 312-329, 2016, Available at SSRN: <https://ssrn.com/abstract=2510013> or <http://dx.doi.org/10.2139/ssrn.2510013>
22. Z. A. Mdaghri, M. El Yadari, A. Benyoussef and A. El Kenz, "Study and analysis of data mining for healthcare," 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016, pp. 77-82, <https://doi.org/10.1109/CIST.2016.7804994>
23. C. Pasupathi and V. Kalavakonda, "Evidence Based health care system using Big Data for disease diagnosis," 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2016, pp. 743-747, <https://doi.org/10.1109/AEEICB.2016.7538393>
24. U. Akhtar, J. W. Lee, H. S. Muhammad Bilal, T. Ali, W. A. Khan and S. Lee, "The Impact of Big Data In Healthcare Analytics," 2020 International Conference on Information Networking (ICOIN), 2020, pp. 61-63, <https://doi.org/10.1109/ICOIN48656.2020.9016588>
25. Benabderrahmane S. (2017) What Can the Big Data Eco-System and Data Analytics Do for E-Health? A Smooth Review Study. In: Rojas I., Ortuño F. (eds) *Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science*, vol 10208. Springer, Cham. https://doi.org/10.1007/978-3-319-56148-6_56
26. Caliebe, A., Leverkus, F., Antes, G. et al. Does big data require a methodological change in medical research? *BMC Med Res Methodol* 19, 125 (2019). <https://doi.org/10.1186/s12874-019-0774-0>
27. Dolley S (2018) Big Data's Role in Precision Public Health. *Front. Public Health* 6:68. <https://doi.org/10.3389/fpubh.2018.00068>
28. L. Elezabeth, V. P. Mishra and J. Dsouza, "The Role of Big Data Mining in Healthcare Applications," 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2018, pp. 256-260, <https://doi.org/10.1109/ICRITO.2018.8748434>

29. P. Halder and I. Pan, "Role of Big Data Analysis in Healthcare Sector: A survey," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 221-225, <https://doi.org/10.1109/ICRCICN.2018.8718684>
30. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform.* 2018;15(3):20170030. Published 2018 May 10. <https://doi.org/10.1515/jib-2017-0030>
31. Wang, Y., et al., Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, *Technol. Forecast. Soc. Change* (2016), <http://dx.doi.org/10.1016/j.techfore.2015.12.019>
32. Adnan K., Akbar R., Khor S.W., Ali A.B.A. (2020) Role and Challenges of Unstructured Big Data in Healthcare. In: Sharma N., Chakrabarti A., Balas V. (eds) *Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing*, vol 1042. Springer, Singapore. https://doi.org/10.1007/978-981-32-9949-8_22
33. Favaretto, M., De Clercq, E. & Elger, B.S. Big Data and discrimination: perils, promises and solutions. A systematic review. *J Big Data* 6, 12 (2019). <https://doi.org/10.1186/s40537-019-0177-4>
34. Hariri, R.H., Fredericks, E.M. & Bowers, K.M. Uncertainty in big data analytics: survey, opportunities, and challenges. *J Big Data* 6, 44 (2019). <https://doi.org/10.1186/s40537-019-0206-3>
35. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract.* 2017;36(1):3-11. <https://doi.org/10.23876/j.krcp.2017.36.1.3>
36. Sokolova M., Matwin S. (2016) Personal Privacy Protection in Time of Big Data. In: Matwin S., Mierniczuk J. (eds) *Challenges in Computational Statistics and Data Mining. Studies in Computational Intelligence*, vol 605. Springer, Cham. https://doi.org/10.1007/978-3-319-18781-5_18
37. A.V. Katin, P.S. Ershov, Yu.E. Khokhlov, S.B. Shaposhnik. Model BD4DE-MM zrelosti raboty s bolshimi dannymi v organizatsii // *Informatsionnoe obshchestvo.* 2021, № 4–5. S. 259–277. https://doi.org/10.52605/16059921_2021_04_259
38. Passport federalnogo proekta «Informatsionnaya infrastruktura». Utverzhden prezidiumom Pravitelstvennoi komissii po tsifrovomu razvitiyu, ispolzovaniyu informatsionnykh tekhnologii dlya uluchsheniya kachestva zhizni i uslovii vedeniya predprinimatelskoi deyatel'nosti (protokol ot 28 maya 2019 g. № 9)
39. Forma № 3-inform. Svedeniya ob ispolzovanii tsifrovyykh tekhnologii i proizvodstve svyazannykh s nimi tovarov i uslug. Utverzhdena Prikazom Rosstat «Ob utverzhdenii formy» ot 30.07.2020 № 424.
40. Rosstat. Itogi federal'nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol'zovanii cifrovyy`x tekhnologij i proizvodstve svyazanny`x s nimi tovarov i uslug» za 2020 god (razmeshheno - 13.10.2021) (accessed on 15.10.2021).

Наука и инновации в информационном обществе

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ ДАНЫМИ В РОССИЙСКОЙ НАУКЕ

Малахов Вадим Александрович

Кандидат исторических наук

Институт истории естествознания и техники им. С. И. Вавилова РАН, центр истории организации науки и науковедения, старший научный сотрудник

Москва, Российская Федерация

yasonbh@mail.ru

Хохлов Юрий Евгеньевич

Кандидат физико-математических наук, доцент

Институт развития информационного общества, председатель совета директоров

РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО

Москва, Российская Федерация

yuri.hohlov@iis.ru

Шапошник Сергей Борисович

Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития, старший научный сотрудник

Петрозаводск, Российская Федерация

sergei.shaposhnik@gmail.com

Аннотация

Разработана концептуальная схема и набор показателей для мониторинга использования технологий работы с большими данными в науке. В концептуальной схеме выделены три основных раздела: уровень зрелости использования технологий в научных организациях, уровень использования технологий в науке, барьеры применения технологий работы с большими данными в науке. Получена оценка уровня использования технологий работы с большими данными в науке и анализ полученных результатов по России в сопоставлении с другими странами. Выявлено, что по сравнению с среднемировым уровнем в России доля статей с использованием технологий работы с большими данными выше в социальных и гуманитарных науках. Выявлена закономерность: публикации, в которых используются большие данные, в среднем цитируются существенно чаще, чем все публикации страны в тех же областях науки.

Ключевые слова

большие данные; библиометрический анализ; патентный анализ; Big Data for Digital Economy; BD4DE; использование больших данных в науке, технологии работы с большими данными; мониторинг

Введение

Сегодня мы живем в эпоху больших данных, когда у человечества появились технологии, с помощью которых можно генерировать, сохранять, обрабатывать и анализировать беспрецедентные до сих пор объемы информации. Большие данные и связанные с ними технологии радикально меняют наше общество, в том числе такую важную сферу деятельности как наука. Появление и распространение больших данных привело к цифровой трансформации исследовательской деятельности: претерпели изменения способы доказательства гипотез, расширился спектр исходной информации для научных исследований и т.д. Возросли требования к квалификации научно-исследовательских кадров – современные научные коллективы трудно себе представить хотя бы без одного специалиста по аналитике больших данных.

© Малахов В.А., Хохлов Ю.Е., Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_200

Развитие новых способов работы с большими данными не только открывает для ученых новый вид ранее недоступных источников, но и способствует появлению широкого спектра новых исследовательских методов и подходов практически в каждой современной научной дисциплине. Так, существенное влияние на Науки о Земле оказывают большие данные, полученные с помощью спутников [1], а в таких науках как социология или психология происходит постепенная переориентация исследователей со сбора данных на их интерпретацию [2; 3]. В настоящее время происходит трансформация целых научных областей, многие традиционные способы сбора и анализа данных уступают место новым методам работы с большими данными [4]. Благодаря таким связанным с большими данными технологиям, как машинное обучение и искусственный интеллект можно утверждать, что сегодня многие знания создаются автономно с помощью инструментов, сделанных учеными, а не непосредственно самими учеными. При этом, по мнению сторонников применения новых технологий и методов работы с большими данными в научных исследованиях, одним из их важных преимуществ является значительная степень свободы полученных результатов от идеологических и теоретических предпочтений конкретных исследователей [5].

Использование методов и технологий работы с большими данными оказывает огромное влияние на все научные дисциплины. Однако что это может означать с точки зрения управления наукой? Для выработки взвешенных управленческих решений как на уровне государства, так и на уровне конкретных организаций (университетов, научных институтов) необходимо не просто констатировать факт влияния больших данных на науку, но и измерить объем этого влияния. Насколько активно используются технологии работы с большими данными в различных областях российской науки? Каковы тренды, связанные с внедрением данных технологий?

Один из подходов к мониторингу и оценке производства, использования и воздействия технологий, решений и услуг для работы с большими данными описан в работе [6] и развивается в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» Национальной технологической инициативы. В данной статье представлена концептуальная схема предметной области мониторинга, посвященная оценке уровня использования технологий работы с большими данными в российской науке. Предложенная схема основана концептуальной схеме BD4DE (Big Data for Digital Economy) [6] и содержит набор показателей, по которым можно оценить влияние больших данных не только на российскую науку в целом, но и отдельно по различным отраслям знания.

1 Обзор литературы

1.1 Библиометрические исследования

Одним из наиболее используемых в современной исследовательской литературе методов для выявления и измерения (мониторинга) тех или иных тенденций в науке является библиометрический подход [7].

В последние годы данный метод начал активно применяться при написании обзоров по исследованиям, связанным с большими данными. Так, поиск научных статей в системе Scopus по ключевым словам "big data" и "bibliometrics" выдает 476 результатов, из которых 412 – статьи, опубликованные за последние 5 лет (с 2016 по 2020 г.). Для поиска в библиографических базах данных платформы Web of Science (далее WoS) статей, посвященных мониторингу и выявлению трендов использования технологий работы с большими данными в различных областях науки, с помощью библиометрии был разработан специальный поисковый образ, включающий в запросе ключевые слова "big data", "bibliometrics", а также названия различных научных дисциплин:

TS=("Big Data" AND ("Physical science*" OR Econometrics OR "Folklore stud*" OR Microbiology OR Virology OR Biochemistry OR Mycology OR Biophysic* OR Genetic* OR botany OR Zoology OR Ornithology OR Entomology OR «Behavioural science*» OR biology OR limnology OR Ecology OR "Chemical science*" OR "Environmental science*" OR "Earth science*" OR Mathematic* OR "Biological science*" OR "Health science*" OR "Medical science*" OR "Agricultural science*" OR "Veterinary science*" OR "Social science*" OR "Educational science*" OR "Political science*" OR Humanities OR "Animal science*" OR "Dairy science*" OR Linguistic* OR "literature stud*" OR "Literary theory" OR Archaeology OR Demography OR Anthropology OR ethnology OR engineering OR Sociology OR geography OR Psychology OR Economics OR chemistry OR physics OR Meteorology OR Geology OR

Mineralogy OR Oceanography OR Astronomy) AND bibliometric*) AND DOP=(2016 OR 2017 OR 2018 OR 2019 OR 2020)

Поиск производился по полю «Тема» (IS), включающего название, аннотацию и ключевые слова статьи, и был ограничен 2016–2020 гг. В общей сложности по данному запросу в WoS было найдено 62 статьи. После изучения аннотаций из 62 статей вручную были отобраны 13 публикаций [8-20], в которых библиометрический подход использован не для подготовки обзорных разделов работ, посвященных исследованиям в области технологий работы с большими данными, а для мониторинга использования технологий работы с большими данными в различных научных областях.

Анализ отобранных в WoS статей показал, что зачастую в работах с применением библиометрического анализа публикаций по тематике больших данных авторы не разделяют статьи, написанные с применением технологий работы с большими данными и статьи, посвященные их разработке, акцентируя внимание на мультидисциплинарности данной области исследований [8]. Авторы статьи [9] применили библиометрический подход для анализа 36 тысяч публикаций 2010-2017 гг., связанных с тематикой больших данных в различных областях науки, и выявили, что если раньше технологии работы с большими данными были представлены практически исключительно в компьютерных науках, то в последние годы большие данные все чаще используются и в других научных областях: медицине, урбанистике, науках об образовании. Библиометрический подход также часто используется для измерения уровня использования технологий работы с большими данными в конкретных научных областях, например в медицинских [10-13], экономических науках [14-17], образовании [18], исследованиях Океана [19] или городского пространства [20].

1.2 Экспертные и социологические опросы

Другим способом оценить уровень применения технологий работы с большими данными в различных областях науки являются социологические и экспертные опросы. Так, одно из крупнейших исследований о цифровизации науки, проведенное экспертами ОЭСР, основано на анализе результатов Международного опроса научных авторов (ISSA2) – глобального онлайн-опроса, разработанного и осуществленного с целью измерения ключевых особенностей цифрового преобразования науки [21]. Данный проект посвящен исследованию потенциального воздействия цифровизации науки на основе ответов почти 12 тыс. авторов научных статей со всего мира. Хотя исследование не ограничено использованием технологий работы с большими данными и освещало также другие вопросы, связанные с цифровизацией науки (например, развитие «Открытой науки»), значительная часть вопросов была посвящена именно большим данным.

На основе ответов респондентов эксперты ОЭСР смогли сформировать целостную картину, описывающую уровень использования технологий работы с большими данными не только в разрезе стран, но и в разрезе отдельных научных дисциплин. При этом вместо того, чтобы создать единый составной индекс цифровизации, эксперты ОЭСР выявили четыре различных ее аспекта в научных исследованиях:

- внедрение цифровых средств с целью повышения производительности на всех этапах научного процесса;
- распространение научных данных с помощью цифровых технологий и расширение доступа к ним;
- использование передовых и ресурсоемких цифровых инструментов для получения информации и разработки прогнозов;
- развитие цифрового самосознания и онлайн-коммуникации в науке.

Для разработки концептуальной схемы мониторинга использования технологий работы с большими данными в науке могут быть также релевантны исследования, посвященные мониторингу уровня цифровизации и внедрения технологий работы с большими данными в других предметных областях. Так, созданная ВОЗ обсерватория электронного (цифрового) здравоохранения раз в несколько лет проводит мониторинг государств-членов этой организации. Для обследования разработан опросный лист, а его заполнение осуществляется в ходе экспертного обсуждения данных обследуемой страны [22]. Схожим образом на основе результатов опроса

экспертов и предприятий, а также данных Евростата специалисты компании IDC по заказу Еврокомиссии провели мониторинг влияния использования технологий, решений и услуг для работы с большими данными на экономику и общество стран ЕС [23]. Специалистами Института статистики ЮНЕСКО была разработана методология и концептуальная схема мониторинга использования ИКТ в системе образования, содержащая более 50 различных статистических показателей [24]. Блок вопросов об использовании аналитики больших данных существует в модельной анкете для бизнеса Евростата [25].

2 Концептуальная схема и показатели мониторинга

2.2 Концептуальная схема мониторинга

На внедрение и использование технологий работы с большими данными в научно-исследовательском процессе влияет множество факторов, среди которых: государственная политика и регулирование работы с большими данными, кадровый потенциал, цифровая инфраструктура, финансирование и пр. Эти характеристики формируют общий уровень зрелости технологий работы с большими данными.

Масштаб и интенсивность использования технологий работы с большими данными в научной сфере, понимание которых крайне полезно для диагностики ситуации на национальном или более локальных уровнях, поддаются измерению про помощи инструментов наукометрии. С точки зрения обеспечения эффективности трансформации национального научно-технологического комплекса высокую значимость приобретает анализ препятствий интеграции технологий для работы с большими данными в исследовательский процесс. Описанные три области оценки образуют комплексный подход к анализу использования технологий работы с большими данными в науке (см. рисунок 1).



Рис. 1. Концептуальная схема мониторинга использования технологий работы с большими данными в науке

Данный подход позволяет получить всестороннее представление о потенциальных точках роста и выделить проблемные зоны распространения технологий работы с большими данными в науке, а также выявить специфику их применения в отдельных областях науки. Весь спектр показателей может рассматриваться в двух разрезах: по областям науки и по секторам деятельности организаций, выполняющих научные исследования и разработки.

2.2 Уровень зрелости использования технологий для работы с большими данными в научных организациях

Первый блок показателей мониторинга посвящен определению уровня зрелости использования технологий работы с большими данными в научных организациях. Оценка уровня зрелости использования технологий работы с большими данными производится по различным аспектам исходя из специально разработанных критериев по пятибалльной шкале (начальный, формирующийся, средний, продвинутый, высокий). Изучение уровня зрелости использования технологий работы с большими данными целесообразно проводить на основе опроса научных и научно-образовательных организаций. Интегральный уровень зрелости образуется при обследовании следующих размерностей [26]:

1. Доля организаций, имеющих продвинутый и выше уровень зрелости работы с большими данными.
2. Доля организаций, имеющих продвинутый и выше уровень зрелости стратегии и регулирования работы с большими данными. Содержательная оценка стратегии работы научной организации с большими данными не может ограничиваться только определением наличия такой стратегии в виде отдельного документа или раздела стратегии цифровой трансформации. Стратегия работы с большими данными как работающий инструмент развития организации должна соответствовать положениям общей стратегии организации, быть обеспечена необходимыми ресурсами и ориентирована на технологические инновации в области работы с большими данными, а также включать план реализации.
3. Доля организаций, имеющих продвинутый и выше уровень кадрового обеспечения работы с большими данными. Человеческий капитал организации, оценивается с точки зрения осведомленности, знаний и компетенций сотрудников организации в области технологий работы с большими данными и их использования при проведении научных исследований.
4. Доля организаций, имеющих продвинутый и выше уровень обеспеченности качественными большими данными. Обеспеченность организации необходимыми для ее деятельности качественными большими данными и эффективная работа с ними являются важными факторами ее успешной деятельности, поэтому эти параметры работы с большими данными были выделены в отдельную подкатегорию.
5. Доля организаций, имеющих продвинутый и выше уровень использования аналитики и инструментов для работы с большими данными. В данной подкатегории объединены два ключевых параметра работы с большими данными в организации: использование специализированных инструментов для работы с большими данными и использование различного вида аналитики больших данных.
6. Доля организаций, имеющих продвинутый и выше уровень зрелости цифровой инфраструктуры и безопасности для работы с большими данными. В данной подкатегории оцениваются два связанных фактора работы с большими данными в организации: инфраструктура работы с большими данными в организации и информационная безопасность работы с большими данными.
7. Доля организаций, имеющих продвинутый и выше уровень зрелости организационных механизмов работы с большими данными. При оценке организационных механизмов работы с большими данными учитываются наличие специального подразделения по работе с большими данными, а также центра компетенций в этой сфере. К институциональным факторам относится также регламентация процессов работы с большими данными.

Блок показателей, характеризующий уровень зрелости использования технологий работы с большими данными, в методологическом отношении не является уникальным для данной предметной области и используется в других отраслевых предметных областях мониторинга использования технологий работы с большими данными. Тем не менее, применительно к научным организациям существует специфика использования этих показателей. В рассматриваемой предметной области нас интересует уровень зрелости использования технологий работы с большими данными именно в преломлении их использования для проведения научно-исследовательских работ, т.е. в данном мониторинге нас не интересует уровень зрелости использования технологий, если организации (в том числе научные

учреждения) используют их не для проведения исследований и разработок, а для других целей. В целом наличие инструмента оценки зрелости использования технологий работы с большими данными в научных организациях поможет выявить «узкие места» и повысить эффективность научно-исследовательской деятельности благодаря внедрению новейших цифровых технологий, включая технологии работы с большими данными.

2.3 Использование технологий для работы с большими данными

В качестве характеристик интенсивности применения в научной сфере новых технологий и методов, как правило, используются и показатели публикационной и патентной активности.

Следует сказать, что мониторинг использования технологий работы с большими данными в науке существенно затруднен задачей отделения случаев непосредственного применения технологий, связанных с большими данными, от их развития и разработки. С одной стороны, учитывая сложность технологий работы с большими данными, новые технологии и их существенные улучшения едва ли могут создаваться без участия специалистов в этой области и являются, как правило, предметом публикаций в журналах, связанных с компьютерными науками и технологиями. С другой стороны, разработка новых методов исследования является частью научной работы и предметом научных публикаций. Можно предположить, что результаты адаптации и апробации в науке существующих технологий и методов (в данном случае – технологий работы с большими данными), которые зачастую связаны с их развитием и доработкой, могут публиковаться в профильных научных журналах (по физике, химии и т.д.).

С учетом этих обстоятельств за основу был принят подход, в рамках которого анализ распространенности использования технологий работы с большими данными в научной деятельности проводится во всех областях науки за исключением компьютерных наук, где концентрируются усилия по разработке новых и развитию существующих технологий. Но при этом для проверки и обоснования такого подхода были проведены эксперименты с выгрузкой и содержательным анализом публикаций в WoS, содержащих в метаописании (ключевых словах, названиях, аннотациях) упоминание технологий работы с большими данными, которые при этом не относятся к компьютерным наукам. Проведенный выборочный анализ публикаций показал применимость такого подхода – в большинстве своем это были работы, основанные на использовании технологий работы с большими данными при проведении исследований.

Если в случае анализа использования технологий работы с большими данными, основанного на библиометрических показателях публикационной активности, указанный подход работает хорошо, то в случае с патентной активностью его применимость, как показали предварительные исследования, ограничена. Это связано с двумя обстоятельствами. Прежде всего, за последние 10 лет было зарегистрировано слишком малое количество российских патентов, созданных с использованием технологий работы с большими данными, чтобы использовать данные об их количестве как один из показателей в концептуальной схеме. Так, поиск в базе данных Федерального института промышленной собственности [27] по названиям и ключевым словам выявил всего 7 патентов, не относящихся к сфере ИКТ. Поиск российских патентов, зарегистрированных по процедуре РСТ, в базе данных Всемирной организации интеллектуальной собственности PatentScore [28] не выявил ни одного патента, не относящегося к сфере ИКТ. Патенты, относящиеся к сфере ИКТ, не учитывались, т.к. предполагается, что в основном это патенты на изобретения, относящиеся непосредственно к области разработки технологий и решений для работы с большими данными, а не на изобретения, для разработки которых использовались технологии работы с большими данными. Использование патентной статистики для измерения использования технологий работы с большими данными как инструмента исследований и разработок осложняется еще и тем обстоятельством, что в патентах, не относящихся к сфере ИКТ, технологии работы с большими данными зачастую используются как составная часть сложной технологии (изобретения), а не как метод проведения исследований и разработок. С учетом этих обстоятельств показатели, связанные с патентной статистикой, не были включены в систему мониторинга.

Схожие проблемы есть и с анализом диссертаций. Поиск в системе ЕГИСУ НИОКТР [291] по диссертациям на соискание ученых степеней, дал только 9 защищенных в России за последние 10 лет диссертаций, в названии и ключевых словах которых упоминались большие данные и основные термины, с ними связанные, но которые не относились к сфере ИКТ.

Поэтому для оценки использования технологий работы с большими данными была разработана система, состоящая из 3-х показателей, отражающих публикационную активность российских авторов:

1. Доля российских публикаций с использованием технологий работы с большими данными, в общем российском публикационном потоке.

Показатель демонстрирует, насколько широко используются технологии работы с большими данными при проведении исследований.

2. Средняя нормализованная цитируемость публикаций с использованием технологий работы с большими данными.

Важную роль в оценке потенциала национальной науки как фактора, влияющего на развитие и использование технологий работы с большими данными, играет не только количество подготовленных работ, но также их качественные характеристики, которые измеряются, в частности, показателями «воздействия», основанными на статистике цитируемости. В данном случае используется показатель средней нормализованной цитируемости. Нормализация осуществляется делением числа цитирований, полученных публикаций, на среднюю по миру цитируемость работ в данной области того же года и типа. Такой подход позволяет при оценке убрать зависимость числа цитирований от года выхода публикации (более ранние в среднем цитируются выше – набирают больше ссылок), типа (например, обзоры цитируются чаще) и области исследований (в более «крупных» областях средняя цитируемость выше). При интерпретации значений данного показателя исходят из следующего: если показатель меньше 1, то цитируемость ниже среднемировой, если больше, то выше.

3. Доля российских публикаций с использованием технологий работы с большими данными, подготовленных в международном соавторстве, относительно всех российских публикаций с использованием технологий работы с большими данными.

Участие в международном научном сотрудничестве является важным каналом обмена знаниями и исследовательскими практиками, что особенно важно для таких новых методов исследования.

2.4 Барьеры применения технологий работы с большими данными

Для эффективного использования технологий работы с большими данными необходимо идентифицировать основные барьеры, препятствующие их применению. Изучение барьеров для внедрения технологий работы с большими данными в научной сфере целесообразно проводить на основе опроса научных и научно-образовательных организаций. Перечень основных барьеров использования технологий работы с большими данными в научных и научно-образовательных учреждениях основывается на методологии проводимых Евростатом опросов по использованию ИКТ на предприятиях [30] и состоит из 9 пунктов:

1. Доля организаций считающих, что затраты на технологии работы с большими данными слишком высокие по сравнению с выгодами.
2. Доля организаций считающих, что у нее недостаточно человеческих ресурсов, знаний, навыков.
3. Доля организаций считающих, что источников больших данных как внутри, так и за пределами предприятия недостаточно для решения нужных задач.
4. Доля организаций считающих, что у нее недостаточно развита ИКТ-инфраструктура.
5. Доля организаций считающих, что законы о персональных данных и их конфиденциальности мешают использованию технологий работы с большими данными.
6. Доля организаций считающих, что технологии работы с большими данными не являются приоритетом для их деятельности.
7. Доля организаций считающих, что источников больших данных для решения текущих задач организации недостаточно.
8. Доля организаций считающих, что технологии работы с большими данными бесполезны для организации.
9. Доля организаций, указавших другие причины.

Мониторинг и анализ существующих барьеров использования технологий работы с большими данными в науке позволит идентифицировать «узкие места», препятствующие

внедрению новых технологий в научно-исследовательской деятельности. Это в свою очередь необходимо для выработки действенных рекомендаций по корректировке стратегии развития науки как на уровне государства, так и на уровне конкретных организаций.

Среди возможных барьеров внедрения новых технологий, предложенных в анкете, фигурируют такие важные препятствия, как ограниченность финансовых ресурсов, нехватка квалифицированных кадров, отсутствие современной цифровой инфраструктуры и пр. В целом показатели, использованные в данном блоке мониторинга, идентичны показателям мониторинга использования технологий работы с большими данными в других предметных областях. Тем не менее, как и в случае с блоком показателей, характеризующих уровень зрелости использования технологий, для научных организаций существует своя специфика, заключающаяся в том, что в данном мониторинге нас интересуют исключительно технологии работы с большими данными, использующиеся при проведении НИОКР.

Содержательный анализ барьеров, упомянутых в заключительном пункте («другие причины»), может способствовать раскрытию препятствий применения технологий работы с большими данными специфических именно для сферы НИОКР, тем самым способствуя генерации профильных инициатив для государственной научно-технической политики.

3 Методология мониторинга

В данной работе осуществлена пилотная реализация мониторинга использования в научной сфере технологий работы с большими данными. Данный раздел мониторинга основывается на наукометрических показателях (количество и цитируемость публикаций, количество публикаций, написанных в международном соавторстве).

Основным источником для подсчета библиометрических показателей является платформа WoS и ее аналитический инструмент InCites. Поиск публикаций российских ученых (в том числе высокочитруемых и в соавторстве с иностранными соавторами), использовавших при проведении исследований технологии больших данных, проводился по WoS Core Collection с помощью специально сформированного поискового образа. Данный поисковый образ был разработан с привлечением экспертизы ведущих российских специалистов в области больших данных [6] и выглядит следующим образом:

TS=(“Big data*” OR Bigdata OR “Large dataset*” OR “Massive Data*” OR “Data science” OR “Data* mining” OR Datamining OR “Text mining” OR Hadoop* OR MapReduce OR “Map Reduce” OR “Unstructured data*” OR “Semistructured data*” OR “Semi-structured data*” OR “Data analytic*” OR “Descriptive analytic*” OR “Diagnostic Analytic*” OR “Predictive Analytic*” OR “Prescriptive Analytic*”) AND DOP=(2019)

Разработанный поисковый образ позволяет максимально широко охватить статьи, связанные с технологиями работы с большими данными. Хронологически поиск был ограничен 2019 г. и проводился по полю «Тема» (TS), включающему в себя название публикации, ключевые слова и аннотацию. Как уже было сказано выше, для того чтобы в полученную выборку вошли статьи, в которых использовались связанные с большими данными технологии, а не статьи, посвященные их разработке, из результата поиска исключались все публикации из двух областей науки по классификатору ОЭСР: «Компьютерные и информационные науки» (Computer and information sciences) и «Электротехника, электронная техника, информационные технологии» (Electrical engineering, electronic engineering).

Полученная выборка (24563 публикации) была проанализирована в разрезе областей наук и стран. Ограничением использованного метода можно считать то, что, несмотря на привлечение при разработке поискового образа экспертизы ведущих специалистов в области технологий работы с большими данными, в полученную выборку могло попасть некоторое количество публикаций, не связанных с использованием технологий работы с большими данными. И, наоборот, за пределами разработанного запроса могло остаться определенное количество публикаций, при написании которых использовались новые технологии. Одна из возможных гипотез, почему разработанный запрос может не охватывать весь массив публикаций с использованием технологий работы с большими данными в научных исследованиях, заключается в том, что при анализе больших данных могут использоваться и упоминаться в публикациях не сами технологии, а основанные на них готовые решения для оперирования большими данными (например, Apache SAMOA, Data Analytics Supercomputer и т.п.). Эта гипотеза была проверена:

использование таких решений без упоминания указанных выше ключевых слов из поискового образа действительно встречается, но количество таких публикаций составляют менее 1% от общего массива публикаций, выявленных по запросу.

Таким образом, можно сказать, что разработанный подход является достаточно точным инструментом выделения массива публикаций российских ученых с использованием технологий работы с большими данными, а неточности в выборке не превышают статистической погрешности.

4 Результаты

4.1 Доля публикаций, связанных с использованием технологий работы с большими данными

Полученные в результате библиометрического анализа данные по доле научных публикаций с использованием технологий работы с большими данными представлены на рисунке 2. Как видно из представленных данных, Россия находится на 21-м месте по этому показателю среди 25 стран с наибольшим числом публикаций с использованием технологий работы с большими данными. Россия опережает такую высокотехнологическую страну как Японию, но отставая от Китая, США, Южной Кореи, Канады и большинства крупных научных держав Европы. Интересно, что лидером по доле публикаций с использованием новых технологий с большим отрывом является Иран. Высокий процент подобных публикаций имеют также ученые из Сингапура и Гонконга.

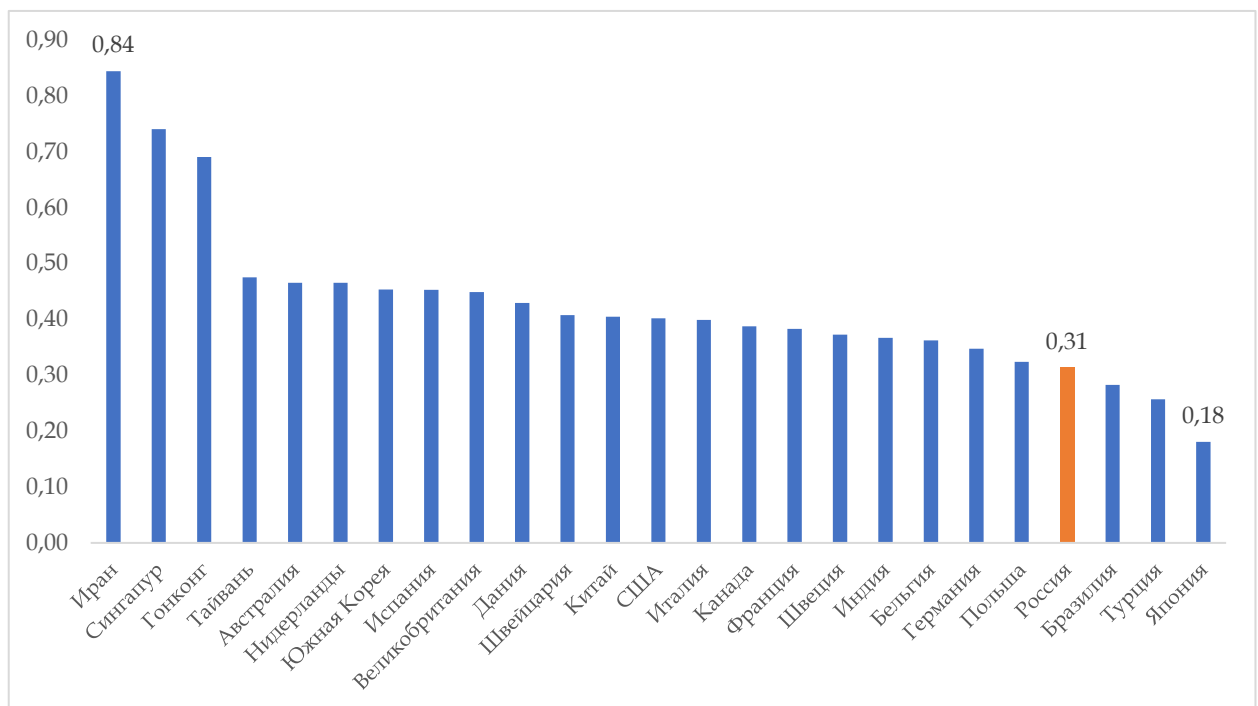


Рисунок 2. Доля публикаций с использованием технологий работы с большими данными (25 стран-лидеров по числу публикаций с большими данными, все области науки кроме *Computer and information sciences* и *Electrical and electronic engineering*, WoS, 2019 г.), %

Рассмотрим более детально использование технологий работы с большими данными в различных областях науки учеными России, мира в среднем, а также Китая и США (рисунок 2). В целом в мире при проведении исследований чаще всего работу с большими данными используют представители естественных (1,12 % от всех публикаций в данной области наук), технических (1,01%) и общественных наук (0,98%), реже – представители медицинских (0,26%), сельскохозяйственных (0,23%) и гуманитарных наук (0,18%). При этом наблюдается интересная национальная специфика: в России и Китае интенсивнее всего технологии работы с большими данными использовались представителями общественных наук, а в США и в среднем по миру лидируют естественные и технические науки, а общественные находятся на третьем месте (рисунок 3).

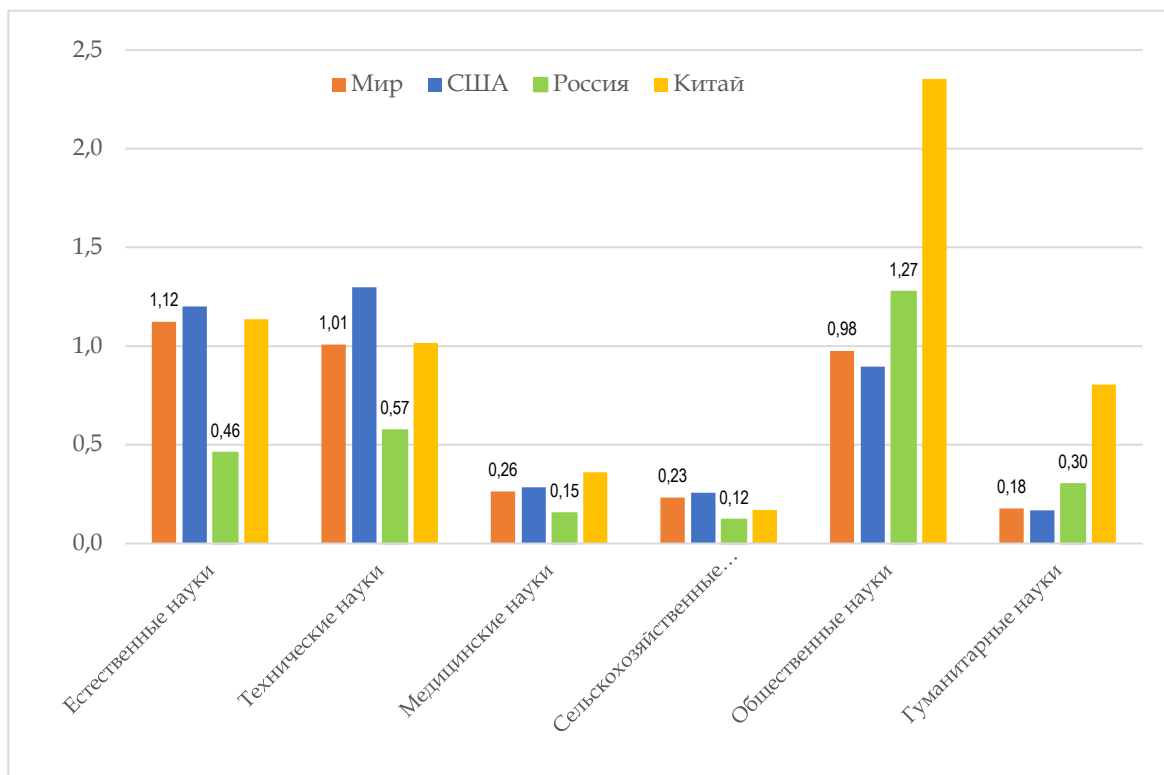


Рисунок 3. Доля публикаций с использованием технологий работы с большими данными, в различных областях науки (WoS, 2019 г.), %

В качестве возможной объяснительной гипотезы более активного использования больших данных в китайских исследованиях в области общественных и гуманитарных наук можно указать на использование в стране таких инструментов контроля, как социальный рейтинг, который предполагает разработку методов анализа поведения людей на основе больших массивов данных, в том числе цифровых следов, оставляемых при использовании электронных сервисов (торговых площадок, социальных сетей, поисковиков и т.д.).

Более детальная картина использования технологий работы с большими данными в различных научных дисциплинах представлена ниже на рисунке 4.

Практически во всех технических и естественнонаучных дисциплинах у российских авторов доля публикаций с использованием технологий работы с большими данными ниже, чем в среднем по миру. В то же время в таких общественных науках как экономика, политология, социология и науки об образовании уровень использования новых технологий выше, чем в среднем по миру в 2-3 раза. Лидерами в данной области являются экономисты – в 2,37% российских публикаций по экономике 2019 г., индексируемых в WoS, были использованы технологии работы с большими данными. Интересно, что в целом по миру большие данные чаще всего используются в работах, посвященных исследованиям СМИ и массовых публикаций (2,62% публикаций), однако в России лишь 1,22% публикаций по данной тематике были написаны с применением новых технологий, что может свидетельствовать о низком уровне развития данной области исследований в России. Среди гуманитарных наук в России выделяется лингвистика, представители которой также относительно активно используют новые методы в своих исследованиях – в среднем 0,55% публикаций были написаны с их применением, в то время как в среднем по миру – лишь 0,15% (рисунок 4). Это можно объяснить традиционно сильной в стране школой математической лингвистики, а также успешным развитием отечественных компаний, активно разрабатывающих и использующих методы компьютерной лингвистики (АВВУУ, Яндекс и др.).

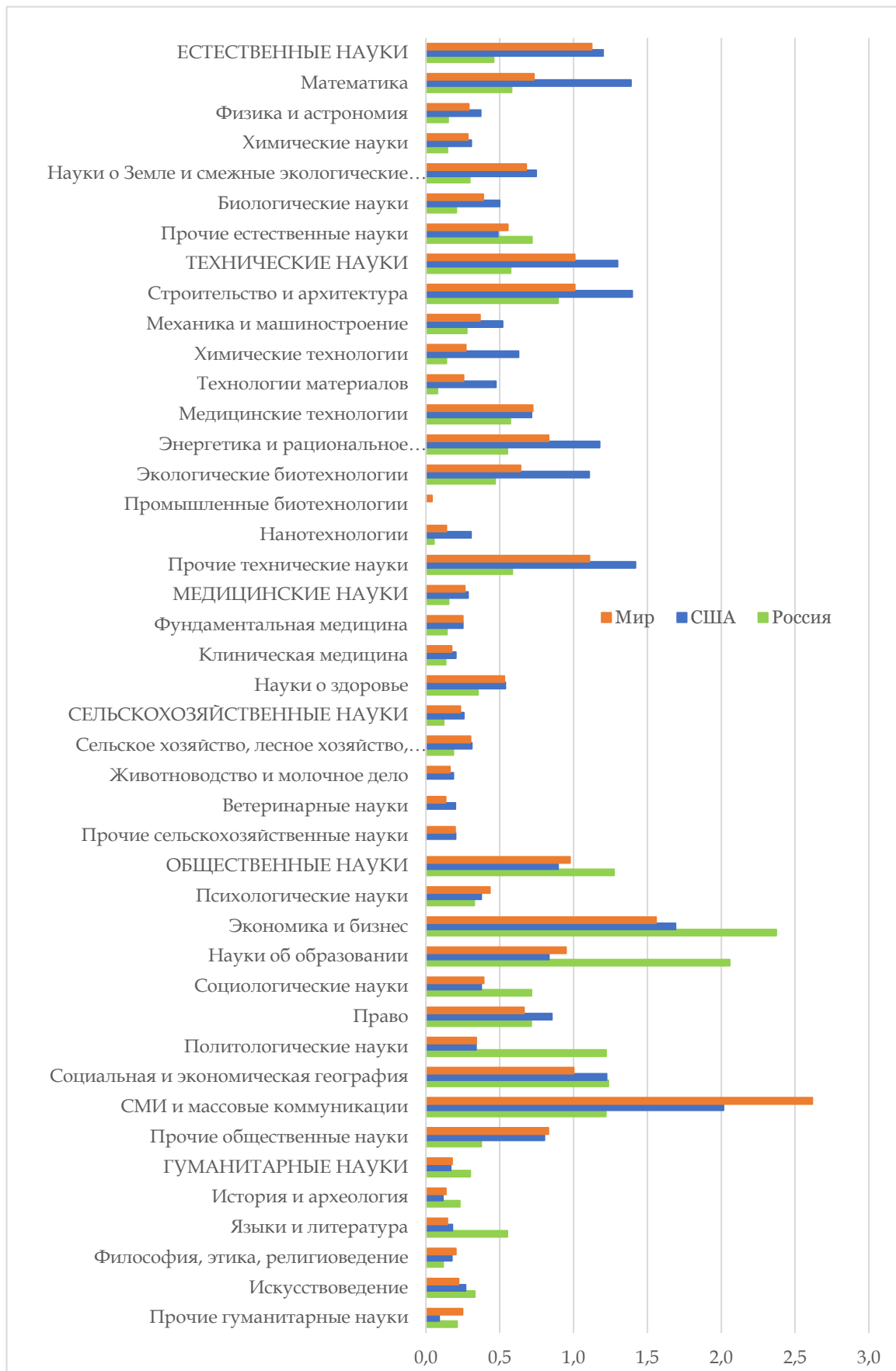


Рисунок 4. Доля публикаций с использованием технологий работы с большими данными, в различных областях науки (WoS, 2019 г.), %

4.2 Цитируемость публикаций, связанных с использованием технологий работы с большими данными

Как цитируются научные публикации с использованием больших данных? На рисунке 5 приведены сравнительные данные по средней нормализованной цитируемости публикаций с использованием больших данных и всех публикаций, индексируемых в WoS (за исключением областей науки, связанных с исследованиями и разработками в области компьютерных наук). Данные приведены для 25 стран, которые имеют наибольшее число публикаций с использованием больших данных. Нормализация цитируемости проводится по году, типу и научной дисциплине, что позволяет корректно проводить сравнения цитируемости, в том числе международные (см. раздел 2.3).

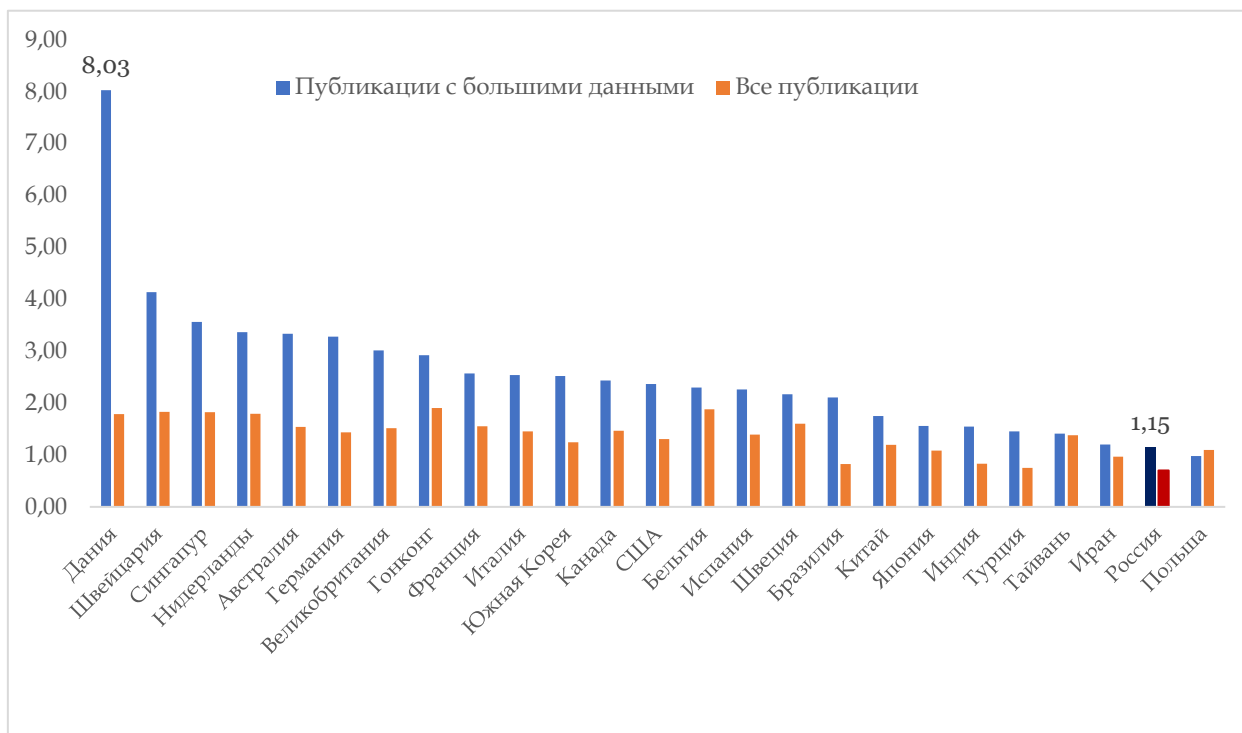


Рисунок 5. Средняя нормализованная цитируемость публикаций с использованием технологий работы с большими данными и всех публикаций (25 стран-лидеров по числу публикаций с большими данными, все области науки, кроме Computer and information sciences и Electrical and electronic engineering, WoS, 2019 г.), %

Россия занимает предпоследнее место среди 25 стран по этому показателю с нормализованной цитируемостью 1,15, что выше среднемировой (общая нормализованная цитируемость российских работ 0,7 – ниже среднемировых показателей).

Интересным результатом, полученным в ходе исследования, является явно выраженная закономерность: публикации, в которых используются большие данные, в среднем цитируются существенно чаще, чем все публикации страны в тех же областях науки. Эта закономерность характерна для всех обследованных стран (кроме Польши). Можно говорить о большем воздействии этих публикаций на мировую науку. Интерес научного сообщества к ним связан с относительно новым методом исследования, который развивается и открывает новые возможности. Работы, основанные на использовании технологий работы с большими данными, формируют быстрорастущее направление исследований.

4.3 Международное сотрудничество в исследованиях, связанных с использованием технологий работы с большими данными

На рисунке 6 приведены данные о международном соавторстве публикаций, связанных с использованием технологий работы с большими данными, для 25 стран-лидеров по числу публикаций с большими данными.

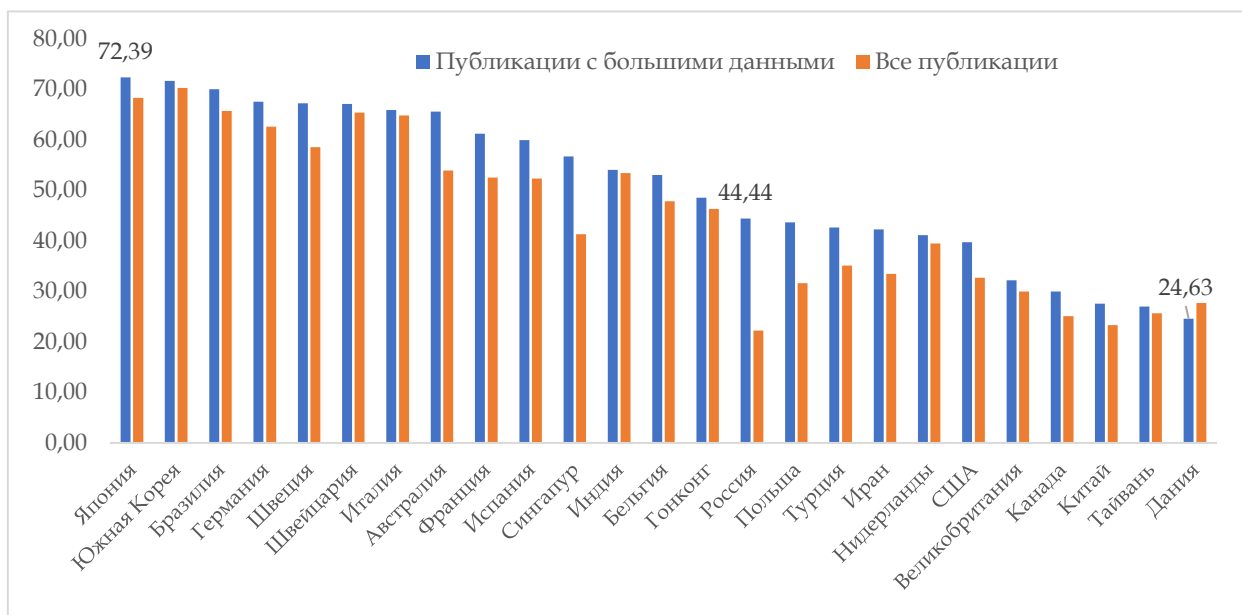


Рисунок 6. Международное соавторство публикаций с использованием технологий работы с большими данными и всех публикаций (25 стран-лидеров по числу публикаций с большими данными, все области науки, кроме Computer and information sciences и Electrical and electronic engineering, WoS, 2019 г.), %

Россия занимает 15-е место по этому показателю среди 25 стран с долей работ с международным авторским коллективом 44,44%. Здесь также можно отметить менее выраженную (чем в случае цитируемости) тенденцию большей вовлеченности в международное научное сотрудничество исследователей, использующих технологии работы с большими данными. Характерно, что по доле всех работ, подготовленных международным авторским коллективом, Россия занимает последнее место среди обследованных 25 стран и у нее разрыв между сотрудничеством в области использования больших данных и общими показателями наиболее существенный (практически в 2 раза – 22,26% для всех публикаций). Дальнейшие наблюдения покажут, насколько этот разрыв и место сохранятся – для России с середины 2000-х характерна стагнация международного научного сотрудничества, идущая вразрез с общемировым тенденциям его расширения [31; 32].

4.4 Сравнение с данными социологических исследований

Интерес представляет сравнение данных, полученных в результате библиометрического анализа, с результатами исследования ОЭСР, опирающегося на международный опрос научных авторов (ISSA2) [21]. По данным ОЭСР, 24% опрошенных авторов научных статей (взвешенная оценка с поправкой на не ответивших респондентов) используют в своих исследованиях (или разрабатывают) технологии работы с большими данными. При этом под большими данными подразумеваются данные с такими характеристиками объема, сложности и неоднородности, которые могут быть обработаны только с помощью нетрадиционных инструментов и методов.

В исследовании все авторы статей были распределены по научным направлениям на основе тематической классификации журналов, в которых они публиковались, используемой в системе Scopus. Чаще всего на использование технологий работы с большими данными указывали представители компьютерных наук (40% респондентов), авторы статей в мультидисциплинарных журналах (34%), генетики, микробиологи и биохимики (30%), представители инженерных наук (30%), в то время как реже всего – гуманитарии (10%) и представители психологических наук (12%) (таблица 1). В целом данные ОЭСР о высоком уровне использования технологий работы с большими данными в технических науках и о низком уровне использования этих технологий в гуманитарных науках и психологии совпадают с полученными нами результатами. Несколько выбиваются из общей картины биологические науки: полученные авторами данные говорят о среднем уровне использования технологий работы с большими данными в этом научном направлении.

Данное несоответствие можно объяснить тем, что в исследовании ISSA2 использовалась более узкая классификация, выделяющая генетику, биохимию, молекулярную биологию и иммунологию в отдельное направление.

Таблица 1. Доля авторов научных статей, использующих или разрабатывающих методы и технологии работы с большими данными (по данным исследования ISSA2)

Научное направление	Доля (%)
Мультидисциплинарные журналы	34
Сельскохозяйственные и биологические науки	25
Искусство и гуманитарные науки	10
Биохимия, генетика, молекулярная биология, иммунология и микробиология	30
Бизнес	18
Химическая инженерия и химия	25
Компьютерные науки	41
Науки о Земле и планетах	26
Экономика, финансы и наука о принятии решений	22
Энергетика и экологические науки	23
Инженерия	30
Материаловедение	28
Математика	21
Медицина и профессии в области здравоохранения	17
Неврология, фармакология, токсикология и фармацевтика	23
Физика и астрономия	22
Социальные науки и психология	12
Всего	24

Источник: [21].

Что касается географического распределения использования учеными технологий работы с большими данными, то по данным исследования ISSA2, в России их используют 16% ученых, в США – 20% ученых (т.е. меньше среднемирового уровня), а лидером в этой области являются исследователи из Индии (40%). Значительно более высокие значения показателей использования технологий работы с большими данными в науке в исследовании ISSA2 по сравнению с полученными нами данными объясняются несколькими причинами. Во-первых, в исследовании ОЭСР оценивалась доля ученых, использующих новые технологии, в то время как мы оцениваем долю исследований (статей), проведенных с использованием этих технологий. Например, если один ученый проводил 10 исследований без использования технологий работы с большими данными и 1 исследование с их использованием, то в статистику ISSA2 он попадал как исследователь, использующий новые технологии. Во-вторых, в исследовании ISSA2 не проведено разделение между учеными, разрабатывающими новые технологии, и учеными, их использующими. В-третьих, исследование ISSA2 основывалось на опросе самих авторов, и они могли достаточно вольно трактовать, что считать работой с большими данными и использованием соответствующих технологий, в то время как в нашем исследовании производится механический (автоматический) отбор статей на основе строгого поискового образа. И, наконец, разницу в результатах и в доле публикаций и доле исследователей, использующих большие данные, можно объяснить особенностью опроса – авторы, которые не используют большие данные и не планируют этого делать, очевидно существенно реже отвечали на анкету, не имея необходимых мотивации и компетенций (возврат заполненных анкет в обследовании составлял менее 9%).

Заключение

Таким образом, нами были разработаны концептуальная схема и набор показателей для мониторинга использования технологий работы с большими данными в науке. По результатам пилотного расчета показателей использования технологий работы с большими данными и сравнительного анализа полученных результатов по России с другими странами можно сделать следующие выводы:

- 1) разработанная методология адекватно отражает предметную область и может быть использована при проведении дальнейших мониторинговых обследований;
- 2) предложенная концептуальная схема мониторинга и система показателей масштабируемы и позволяют включать в систему мониторинга (без изменения ее концептуальных рамок) новые показатели;
- 3) опыт разработки и реализации системы мониторинга использования технологий работы с большими данными в науке может быть распространен на другие сквозные технологии цифровой экономики.

Полученные результаты позволили выявить некоторые слабые и сильные стороны. В большинстве научных областей уровень использования технологий работы с большими данными у российских ученых ниже среднемирового. Это особенно характерно для естественнонаучных и технических дисциплин, в которых данные технологии традиционно используются чаще всего. Это говорит о некотором отставании российской науки в данных областях и о значительном потенциале роста. В то же время в ряде областей (экономические и политические науки, науки об образовании, лингвистика) уровень использования технологий работы с большими данными в российской науке значительно превосходит среднемировую, что свидетельствует о быстром развитии данных исследовательских направлений в России и восприимчивости представляющих их российских ученых к новым технологическим веяниям.

К другим важным результатам пилотного мониторинга использования технологий работы с большими данными в науке можно отнести выявление следующих закономерностей: во-первых, публикации, в которых используются большие данные, в среднем цитируются существенно чаще, чем все публикации страны в тех же областях науки; во-вторых, публикации с использованием технологий работы с большими данными чаще по сравнению с другими публикациями были написаны международными авторскими коллективами. Данные закономерности характерны для абсолютного большинства обследуемых стран и научных областей.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. Guo H., Wang L., Liang D. Big Earth Data from space: a new engine for Earth science // Science Bulletin. 2016. Vol. 61. № 7. P. 505–513. <https://doi.org/10.1007/s11434-016-1041-y>.
2. Губа К. Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. 2018. Т. 17. № 1. С. 213–236. <https://doi.org/10.17323/1728-192X-2018-1-213-236>.
3. Albritton B. H., Tonidandel S. How Can Big Data Science Transform the Psychological Sciences? // The Spanish Journal of Psychology. 2020. № 23. P. 1–5. <https://doi.org/10.1017/SJP.2020.45>.

4. Одинцов А. В. Социология общественного мнения и вызов Big Data // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 3. С. 30–43. <https://doi.org/10.14515/monitoring.2017.3.04>.
5. Sætra H. S. Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science // Integrative Psychological and Behavioral Science. 2018. Vol. 52. № 4. P. 508–522. <https://doi.org/10.1007/s12124-018-9447-5>.
6. Ершова Т.В., Хохлов Ю.Е., Шапошник С.Б. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество, 2021, № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
7. Bellis, N.D. Bibliometrics and Citation Analysis: from the Science Citation Index to Cybermetrics. Lanham: Scarecrow Press, 2009. 417 p.
8. Hu J., Zhang Y. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization // Scientometrics. 2017. Vol. 112. № 1. P. 91–109. <https://doi.org/10.1007/s11192-017-2383-1>
9. Mohammadi E., Karami A. Exploring research trends in big data across disciplines: A text mining analysis // Journal of Information Science. – 2020. – С. 0165551520932855. <https://doi.org/10.1177/0165551520932855>
10. Liao H. et al. A Bibliometric Analysis and Visualization of Medical Big Data Research // Sustainability. 2018. Vol. 10. № 1. P. 166. DOI: <https://doi.org/10.3390/su10010166>
11. Galetsi P., Katsaliaki K. Big data analytics in health: an overview and bibliometric study of research activity // Health Information & Libraries Journal. 2020. Vol. 37. № 1. P. 5–25. <https://doi.org/10.1111/hir.12286>
12. Galetsi P., Katsaliaki K. A review of the literature on big data analytics in healthcare // Journal of the Operational Research Society. 2020. Vol. 71. № 10. P. 1511–1529. <https://doi.org/10.1080/01605682.2019.1630328>
13. Gu D. et al. Visualizing the knowledge structure and evolution of big data research in healthcare informatics // International journal of medical informatics. 2017. Vol. 98. P. 22–32. <https://doi.org/10.1016/j.ijmedinf.2016.11.006>
14. Zhang Y. et al. A bibliometric review of a decade of research: Big data in business research—Setting a research agenda // Journal of Business Research. 2020. <https://doi.org/10.1016/j.jbusres.2020.11.004>
15. Somanathan A. R., Rama S. K. A Bibliometric Review of Stock Market Prediction: Perspective of Emerging Markets // Applied Computer Systems. 2020. Vol. 25. № 2. P. 77–86. <https://doi.org/10.2478/acss-2020-0010>
16. Guo L. An Overview of Research on Big Data in Economics Based on Bibliometric Software Citespace and CSSCI Database // 3rd International Conference on Judicial, Administrative and Humanitarian Problems of State Structures and Economic Subjects (JAHP 2018). Atlantis Press, 2018. P. 177–181.
17. López-Robles J. R. et al. The last five years of Big Data Research in Economics, Econometrics and Finance: Identification and conceptual analysis // Procedia computer science. 2019. Vol. 162. P. 729–736.
18. Wang W. Analysis of Research Hotspots on Big Data in Education: From Bibliometrics Perspective // 2016 International Conference on Management Science and Innovative Education. – Atlantis Press, 2016. P. 14–19. <https://doi.org/10.2991/msie-16.2016.4>
19. Wu J. et al. Development Trends and Frontiers of Ocean Big Data Research Based on CiteSpace // Water. 2020. Vol. 12. №.6. P. 1560. <https://doi.org/10.3390/w12061560>
20. González-Zamar M. D. et al. IoT technology applications-based smart cities: Research analysis // Electronics. 2020. Vol. 9. № 8. P. 1246. <https://doi.org/10.3390/electronics9081246>
21. Bello M., Galindo-Rueda F. Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). OECD Science, Technology and Industry Working Papers, No. 2020/03. Paris: OECD Publishing, 2020. <https://doi.org/10.1787/1b06c47c-en>
22. World Health Organization. Global Observatory for eHealth [Электронный ресурс]. URL: <https://www.who.int/goe/en/> (дата обращения: 08.04.2021).
23. Cattaneo G, Micheletti G., Glennon M, La Croce C., Mitta C. The European Data Market Monitoring Tool. Key facts & figures, first policy conclusions, data landscape and quantified stories. Final study report. Luxembourg: Publications Office of the European Union, 2020. URL:

- <https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update> (дата обращения: 31.03.2021)
24. UNESCO-UIS. Guide to Measuring Information and Communication Technologies (ICT) in Education. Montreal: UNESCO Institute for Statistics. 2009. 139 p. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (дата обращения: 20.04.2021).
 25. Eurostat. Digital economy and society. Data. Database. URL: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_eb_bd&lang=en (дата обращения: 20.04.2021).
 26. Катин А.В., Ершов П.С., Хохлов Ю.Е., Шапошник С.Б. Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, № 4–5. С. 259–277. https://doi.org/10.52605/16059921_2021_04_259
 27. Федеральный институт промышленной собственности. Информационно-поисковая система. URL: <https://www.fips.ru/iiss/> (дата обращения: 20.04.2021).
 28. WIPO. PatenetScope. [Электронный ресурс]. URL: <https://patentscope.wipo.int/search/ru/search.jsf> (дата обращения: 20.04.2021).
 29. ЕГИСУ НИОКТР. URL: <https://www.rosrid.ru/> (дата обращения: 20.04.2021).
 30. Eurostat. Glossary: Community survey on ICT usage in enterprises. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Community_survey_on_ICT_usage_in_enterprises (дата обращения: 20.04.2021).
 31. Шапошник С.Б. Международное научное сотрудничество и публикационная активность российских ученых в Computer Science в 1993–2017 гг.: междисциплинарный и межстрановой сравнительный анализ // Информационное общество, 2018, №6, С. 39–45.
 32. Шапошник С.Б. Сравнительный анализ динамики публикационной активности и международного научного сотрудничества российских ученых в общественных и гуманитарных науках в постсоветский период / Институт истории естествознания и техники им. С. И. Вавилова. Годичная научная конференция, 2019. С. 560–563.

THE USE OF BIG DATA TECHNOLOGIES IN RUSSIAN SCIENCE

Malakhov, Vadim Aleksandrovich

Candidate of historical sciences

Vavilov Institute for the History of Science and Technology of the Russian Academy of Sciences, Center for the history of science organization and scientific research, senior researcher

Moscow, Russia

yasonbh@mail.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor

Moscow, Russia

yuri.hohlov@iis.ru

Shaposhnik, Sergei Borisovich

Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher

Petrozavodsk, Russian Federation

sergei.shaposhnik@gmail.com

Abstract

A conceptual framework and a set of indicators have been developed for monitoring the use of big data technologies in science. The conceptual framework has three main sections: the level of maturity of technology use in scientific organizations, the level of technology use in science, and barriers to the use of big data technologies in science. An assessment of the level of big data technology use in science and a comparative analysis of the results obtained for Russia and other countries were made. It was found that, compared to the global average, Russia has a higher share of articles using big data technologies in the social sciences and humanities. A regularity has been revealed: publications that use big data are cited significantly more frequently on average than all publications of the country in the same fields of science.

Keywords

big data; bibliometric analysis; patent analysis; BD4DE; Big Data for Digital Economy; use of big data in science, big data technologies; monitoring

References

1. Guo H., Wang L., Liang D. Big Earth Data from space: a new engine for Earth science // *Science Bulletin*. 2016. Vol. 61. № 7. P. 505–513. <https://doi.org/10.1007/s11434-016-1041-y>.
2. Guba K. Bol'shie dannye v sociologii: novye dannye, novaja sociologija? // *Sociologicheskoe obozrenie*. 2018. T. 17. № 1. P. 213–236. <https://doi.org/10.17323/1728-192X-2018-1-213-236>.
3. Albritton B. H., Tonidandel S. How Can Big Data Science Transform the Psychological Sciences? // *The Spanish Journal of Psychology*. 2020. № 23. P. 1–5. DOI: <https://doi.org/10.1017/SJP.2020.45>.
4. Odincov A. V. Sociologija obshhestvennogo mneniya i vyzov Big Data // *Monitoring obshhestvennogo mneniya: Jekonomicheskie i social'nye peremeny*. 2017. № 3. P. 30–43. <https://doi.org/10.14515/monitoring.2017.3.04>.
5. Sætra H. S. Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science // *Integrative Psychological and Behavioral Science*. 2018. Vol. 52. № 4. P. 508–522. <https://doi.org/10.1007/s12124-018-9447-5>.
6. Ershova T.V., Hohlov Yu.E., Shaposhnik S.B. Metodologija monitoringa razvitija i ispol'zovaniya tehnologij raboty s bol'shimi dannymi // *Informacionnoe obshhestvo, 2021, № 4–5. S. 2–32*. https://doi.org/10.52605/16059921_2021_04_02

7. Bellis, N.D. *Bibliometrics and Citation Analysis: from the Science Citation Index to Cybermetrics*. Lanham: Scarecrow Press, 2009. 417 p.
8. Hu J., Zhang Y. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization // *Scientometrics*. 2017. Vol. 112. № 1. P. 91-109. <https://doi.org/10.1007/s11192-017-2383-1>
9. Mohammadi E., Karami A. Exploring research trends in big data across disciplines: A text mining analysis // *Journal of Information Science*. – 2020. – C. 0165551520932855. <https://doi.org/10.1177/0165551520932855>
10. Liao H. et al. A Bibliometric Analysis and Visualization of Medical Big Data Research // *Sustainability*. 2018. Vol. 10. № 1. P. 166. DOI: <https://doi.org/10.3390/su10010166>
11. Galetsi P., Katsaliaki K. Big data analytics in health: an overview and bibliometric study of research activity // *Health Information & Libraries Journal*. 2020. Vol. 37. № 1. P. 5–25. <https://doi.org/10.1111/hir.12286>
12. Galetsi P., Katsaliaki K. A review of the literature on big data analytics in healthcare // *Journal of the Operational Research Society*. 2020. Vol. 71. № 10. P. 1511–1529. <https://doi.org/10.1080/01605682.2019.1630328>
13. Gu D. et al. Visualizing the knowledge structure and evolution of big data research in healthcare informatics // *International journal of medical informatics*. 2017. Vol. 98. P. 22–32. <https://doi.org/10.1016/j.ijmedinf.2016.11.006>
14. Zhang Y. et al. A bibliometric review of a decade of research: Big data in business research—Setting a research agenda // *Journal of Business Research*. 2020. <https://doi.org/10.1016/j.jbusres.2020.11.004>
15. Somanathan A. R., Rama S. K. A Bibliometric Review of Stock Market Prediction: Perspective of Emerging Markets // *Applied Computer Systems*. 2020. Vol. 25. № 2. P. 77-86. <https://doi.org/10.2478/acss-2020-0010>
16. Guo L. An Overview of Research on Big Data in Economics Based on Bibliometric Software Citespace and CSSCI Database // *3rd International Conference on Judicial, Administrative and Humanitarian Problems of State Structures and Economic Subjects (JAHP 2018)*. Atlantis Press, 2018. P. 177-181.
17. López-Robles J. R. et al. The last five years of Big Data Research in Economics, Econometrics and Finance: Identification and conceptual analysis // *Procedia computer science*. 2019. Vol. 162. P. 729-736.
18. Wang W. Analysis of Research Hotspots on Big Data in Education: From Bibliometrics Perspective // *2016 International Conference on Management Science and Innovative Education*. – Atlantis Press, 2016. P. 14–19. <https://doi.org/10.2991/msie-16.2016.4>
19. Wu J. et al. Development Trends and Frontiers of Ocean Big Data Research Based on CiteSpace // *Water*. 2020. Vol. 12. №.6. P. 1560. <https://doi.org/10.3390/w12061560>
20. González-Zamar M. D. et al. IoT technology applications-based smart cities: Research analysis // *Electronics*. 2020. Vol. 9. № 8. P. 1246. <https://doi.org/10.3390/electronics9081246>
21. Bello M., Galindo-Rueda F. Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). OECD Science, Technology and Industry Working Papers, No. 2020/03. Paris: OECD Publishing, 2020. <https://doi.org/10.1787/1b06c47c-en>
22. World Health Organization. Global Observatory for eHealth. URL: <https://www.who.int/goe/en/> (accessed on 08.04.2021).
23. Cattaneo G, Micheletti G., Glennon M, La Croce C., Mitta C. The European Data Market Monitoring Tool. Key facts & figures, first policy conclusions, data landscape and quantified stories. Final study report. Luxembourg: Publications Office of the European Union, 2020. URL: <https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update> (accessed on 31.03.2021)
24. UNESCO-UIS. Guide to Measuring Information and Communication Technologies (ICT) in Education. Montreal: UNESCO Institute for Statistics. 2009. 139 p. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf (accessed on 20.04.2021).
25. Eurostat. Digital economy and society. Data. Database. URL: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_eb_bd&lang=en (accessed on 20.04.2021).

26. Ershov P.S., Katin A.V., Hohlov Yu.E., Shaposhnik S.B. Model' BD4DE-MM zrelosti raboty s bol'shimi dannymi v organizacii // Informacionnoe obshhestvo, 2021, № 4–5. S. 259–277. https://doi.org/10.52605/16059921_2021_04_259
27. Federal Institute of Industrial Property. Information and retrieval system. URL: <https://www.fips.ru/iiss/> (accessed on 20.04.2021)..
28. WIPO. PatenetScope. URL: <https://patentscope.wipo.int/search/ru/search.jsf> (accessed on 20.04.2021).
29. Unified state information system of accounting research, development and technological civil works. URL: <https://www.rosrid.ru/> (accessed on 20.04.2021).
30. Eurostat. Glossary: Community survey on ICT usage in enterprises. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Community_survey_on_ICT_usage_in_enterprises (accessed on 20.04.2021).
31. Shaposhnik S.B. Mezhdunarodnoe nauchnoe sotrudnichestvo i publikacionnaja aktivnost' rossijskih uchenyh v Computer Science v 1993-2017 gg.: mezhdisciplinarnyj i mezhsranovoj sravnitel'nyj analiz // Informacionnoe obshhestvo, 2018, № 6, P. 39-45.
32. Shaposhnik S.B. Sravnitel'nyj analiz dinamiki publikacionnoj aktivnosti i mezhdunarodnogo nauchnogo sotrudnichestva rossijskih uchenyh v obshhestvennyh i gumanitarnyh naukah v postsovetskij period / Institut istorii estestvoznaniya i tehniki im. S. I.Vavilova. Godichnaja nauchnaja konferencija, 2019. P. 560-563.

Информационное общество и право**СТАНДАРТИЗАЦИЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ:
МЕЖДУНАРОДНЫЕ И НАЦИОНАЛЬНЫЕ СТАНДАРТЫ****Аверкин Алексей Николаевич**

*Кандидат физико-математических наук, доцент
Российский экономический университет им. Г.В. Плеханова, Учебно-научная лаборатория искусственного интеллекта, нейротехнологий и бизнес-аналитики, ведущий научный сотрудник
Москва, Российская Федерация
averkin2003@inbox.ru*

Афанасьев Сергей Дмитриевич

*Кандидат юридических наук
МГУ имени М.В. Ломоносова, Национальный центр цифровой экономики, ведущий специалист
Московский государственный областной университет, Институт экономики, управления и права, юридический факультет, кафедра конституционного и муниципального права, доцент
Москва, Российская Федерация
sergei.afanasev@digital.msu.ru*

Микрюков Андрей Александрович

*Кандидат технических наук, доцент
Российский экономический университет им. Г.В. Плеханова, Институт математики, информационных систем и цифровой экономики, Кафедра прикладной информатики и информационной безопасности, доцент
Москва, Российская Федерация
mikrakov.aa@rea.ru*

Паджев Валентин Валентинович

*Институт развития информационного общества, руководитель Дирекции правовых программ
Москва, Российская Федерация
vpadzhev@iis.ru*

Райков Александр Николаевич

*Доктор технических наук, профессор
Институт проблем управления имени В.А. Трапезникова РАН, ведущий научный сотрудник
МГУ имени М.В. Ломоносова, Национальный центр цифровой экономики, руководитель департамента интеллектуальных технологий
Москва, Российская Федерация
Alexander.N.Raikov@gmail.com*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель Совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Храмцовская Наталья Александровна

*Кандидат исторических наук
ООО «Электронные офисные системы», ведущий эксперт по управлению документацией
Москва, Российская Федерация
sspchram@tochka.ru*

© Аверкин А.Н., Афанасьев С.Д., Микрюков А.А., Паджев В.В., Райков А.Н., Хохлов Ю.Е., Храмцовская Н.А., 2021.
Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_220

Аннотация

Стандарты, устанавливающие требования к технологиям работы с большими данными, призваны повысить эффективность использования этих технологий в различных отраслях экономики. В статье исследуются особенности международных и российских национальных общеотраслевых стандартов в области данных, основанных на стандартах ISO/IEC 20546, серий 20547-X, 9000 и проектов серии 5259-X, в сфере основных понятий, эталонной архитектуры и требований к большим данным и их качеству, а также подходы к стандартизации требований заказчика к действиям, связанным с использованием больших данных. Делается вывод, что разработка документов по стандартизации больших данных в России имеет большое значение и влечёт необходимость наращивания темпов стандартизации.

Ключевые слова

большие данные; данные; стандартизация; национальный стандарт; международный стандарт; ГОСТ; ISO; Росстандарт; эталонная архитектура; качество данных; искусственный интеллект

Введение

«Большие данные» рассматриваются в качестве одной из ключевых технологий, на которые опирается быстро набирающее популярность направление искусственного интеллекта. Хотя споры о том, как наилучшим образом определить понятие «большие данные», по-прежнему продолжаются, к настоящему времени сложился определенный международный консенсус на основе определения, предложенного в международном стандарте ISO/IEC 20546:2019 [1], адаптированном теперь в России.

В ноябре 2021 года вступил в силу национальный стандарт ГОСТ Р ИСО/МЭК 20546–2021 «Информационные технологии. Большие данные. Обзор и словарь» [2], в котором закрепляется следующее определение больших данных (п. 3.1.2):

Большие данные (big data): Большие массивы данных, отличающиеся главным образом такими характеристиками, как объем, разнообразие, скорость обработки и/или вариативность, которые требуют использования технологии масштабирования для эффективного хранения, обработки, управления и анализа

Важным уточнением данного термина является содержащееся в примечании к пункту 3.1.2 положение о том, что «термин «большие данные» широко применяется в различных значениях, например, в качестве названия технологии масштабирования, используемой для обработки больших массивов данных».

В готовящемся к публикации международном стандарте по концепциям и терминологии искусственного интеллекта ISO/IEC 22989 [3], в частности, отмечается следующая связь больших данных с системами искусственного интеллекта.

«У больших данных имеется множество применений, в том числе в системах искусственного интеллекта (ИИ). Большие данные делают возможными многие ИИ-системы. Наличие больших коллекций неструктурированных данных в различных областях применения способствует получению новых идей и знаний в результате использования таких методов ИИ, как выявление знаний (knowledge discovery) и распознавание закономерностей (pattern recognition). Доступность огромных объёмов данных для обучения приводит к появлению улучшенных моделей машинного обучения, способных решать задачи в широком спектре приложений» (п. 8.5.1).

Технологии работы с большими данными позволяют обрабатывать большие объёмы данных по сравнению со «стандартными», работать с потоками быстро поступающих данных в очень больших объёмах, а также обрабатывать как структурированные, так и неструктурированные или слабо структурированные данные. Такие особенности данного класса цифровых технологий позволяют выявить скрытые закономерности, ускользающие от ограниченного человеческого восприятия. Это даёт беспрецедентные возможности оптимизации различных сфер деятельности: от системы государственного управления, медицины, образования, финансов или транспорта до промышленного производства или телекоммуникаций.

Большие данные представляют собой, по сути, метод «последней надежды», к которому прибегают там, где невозможны прямые измерения и возникает необходимость опираться на косвенные данные больших объёмов и разнообразия. К таким ситуациям, главным образом,

относятся продвинутое инновационные технологии обработки данных с целью выявления неочевидных закономерностей, часто в ситуациях, когда технические особенности потоков данных (объёмы, скорость, разнообразие и т.д.) таковы, что традиционные, давно разработанные методы обработки «небольших» массивов, как правило, структурированных данных с решением такой задачи не справляются. Однако сводить большие данные только к этим техническим особенностям не совсем верно, поскольку даже за последние десятилетия на наших глазах представления о том, что такое большой объём, изменились многократно – от килобайта в середине 1980-х годов до экзбайтов в настоящее время, и эти изменения в восприятии продолжают. Вследствие такого изменения восприятия многие технологии, которые в прошлом веке относили к большим данным, сегодня перестали считаться таковыми, поскольку они стали общедоступными и широко распространёнными.

Новые проблемы и растущая вычислительная мощность будут стимулировать разработку новых аналитических методов. Существует также потребность в постоянных инновациях в технологиях и методах, которые помогут отдельным лицам и организациям интегрировать, анализировать, визуализировать и потреблять растущий поток больших данных. Таким образом, использование технологий работы с большими данными в современном мире имеет первостепенное значение. Уровень развития и использования информационных технологий в государственном управлении, бизнесе, социальной сфере, научных исследованиях и иных сферах жизни общества достиг высокого уровня зрелости, что влечёт за собой необходимость выработки стандартизированных подходов к применению таких технологий, включая технологии работы с большими данными.

Современное состояние развития технологий работы с большими данными характеризуется отсутствием устоявшихся подходов не только к определению и основным характеристикам самих больших данных, но и к эталонной архитектуре систем для работы с большими данными. Значительная часть специалистов, если они не являются узкопрофильными специалистами по работе с данными, продолжает понимать под большими данными такие наборы данных, которые отличаются большим объёмом, разнообразием и высокой скоростью обработки, в то время как понятие «большие данные» представляет собой намного более многосложную конструкцию. Вряд ли можно себе представить, например, что данные о температуре тела больного, получаемые миллиард раз в секунду, представляют собой именно большие данные как они должны пониматься для грамотного использования и эффективного применения. Научные работники и профессиональные инженеры хорошо понимают, что когда есть выбор между целевым и контролируемым образом собранными «под задачу» «малыми данными» и обработкой, скрывающейся за красивым титулом «больших данных», – этот выбор очевиден и это не большие данные.

Полезно также вспомнить, что технологии работы с большими данными первоначально были разработаны разведывательными службами в середине прошлого века и использовались для получения сведений, которые по тем или иным причинам нельзя было собрать средствами агентурной разведки и дистанционного зондирования – например, посредством сбора и анализа информации из массовых открытых источников. Корректное понимание больших данных и их характеристик актуализируется при разработке соответствующих информационных систем и технологий, которые должны быть совместимы между собой. Устаревшие системы и несовместимые форматы часто препятствуют интеграции данных и основанной на них аналитике [4], которая и является основной целью и ценностью использования технологий работы с большими данными. Кроме того, как и для любой инновации важно понять, какую отдачу, при каких обстоятельствах и затратах технологии работы с большими данными способны обеспечить, какие риски с ними связаны. Решения об использовании или неиспользовании больших данных желательно принимать на основе баланса отдачи, затрат и рисков.

Решение соответствующих проблем заключается, в том числе, в стандартизации технологий, что позволит обеспечить единообразие подходов к пониманию и применению технологий больших данных. В процессе стандартизации устанавливаются общие характеристики, правила и принципы в отношении объекта стандартизации, что направлено на достижение добровольной и многократно применяемой упорядоченности соответствующих объектов. Целями стандартизации технологий работы с большими данными являются содействие социально-экономическому развитию, а также содействие интеграции России в мировую экономику и в международные системы стандартизации в качестве равноправного партнёра.

Предметом данной статьи является описание и анализ процессов стандартизации технологий работы с большими данными на международном и национальном уровнях. Здесь также приводятся: краткая история возникновения и развития тематики больших данных, принятые и разрабатываемые стандарты в области работы с большими данными в областях: терминология и направления стандартизации больших данных в целом, включая взаимосвязь со смежными стандартами для других информационных технологий; варианты использования больших данных; эталонная архитектура больших данных, её концептуальная схема и обеспечение безопасности; стандартизация требований заказчика к действиям, связанным с оперированием большими данными; обеспечение качества данных.

1 Краткая история стандартизации больших данных

Активная деятельность по разработке и внедрению национальных и международных стандартов в области больших данных началась с 2013 г., в первую очередь – Национальным институтом стандартов и технологий США (NIST), в рамках которого в июне 2013 г. создаётся рабочая группа по большим данным для стандартизации эталонной архитектуры больших данных, в которую входят представители основных заинтересованных сторон – бизнеса, власти и научно-образовательного сообщества. В 2015 г. NIST принимает серию стандартов в сфере терминологии, архитектуры больших данных, безопасности и конфиденциальности персональных данных при использовании соответствующих технологий [5], которые были пересмотрены в 2018 [6] и 2019 [7] гг.

С развитием соответствующих технологий было признано необходимым гармонизировать разнообразные подходы к пониманию и применению технологий больших данных, чему способствовала разработка и принятие ряда международных стандартов в рамках объединенного технического комитета Международной организации по стандартизации и Международной электротехнической комиссии (ISO/IEC; ИСО/МЭК), а также координация действий ИСО/МЭК с Международным союзом электросвязи (ITU; МСЭ). Стандартизацию в области больших данных на международном уровне можно условно разделить на два исторических этапа.

Первый этап – разработка и утверждение основополагающих международных стандартов. В ноябре 2013 г. Международный союз электросвязи опубликовал аналитический доклад «Большие данные: сегодня большие, завтра нормальные» [8] в котором анализировался феномен больших данных, нашедших применение сразу в нескольких отраслях экономики, в том числе в секторе телекоммуникаций. В начале 2015 г. был опубликован предварительный доклад Объединённого технического комитета № 1 ИСО/МЭК (ISO/IEC JTC 1) «Большие данные» [9], в котором описаны проблемы и направления дальнейшей международной стандартизации технологий работы с большими данными, представлена оценка состояния требований рынка стандартизации больших данных и предложены приоритеты стандартизации. Доклад в целом стал основой плана работ Подкомитета SC42 «Искусственный интеллект» (ISO/IEC JTC 1/SC 42 Artificial intelligence) и его рабочей группы «Большие данные» (ISO/IEC JTC 1/SC 42/WG 2 Big data), переименованной позднее в рабочую группу «Данные» (ISO/IEC JTC 1/SC 42/WG 2 Data).

В ноябре 2015 г. МСЭ утвердил первый международный стандарт в области больших данных «Большие данные – Требования на основе облачных вычислений и их возможности» [10], а уже в июле 2016 г. принимается дорожная карта стандартизации больших данных [11]. Вслед за этим январе-феврале 2018 г. утверждаются два первых международных стандарта ИСО/МЭК из пятикомпонентной серии стандартов ИСО/МЭК 20547-Х, посвящённой стандартизации эталонной архитектуры больших данных: часть 2 «Варианты использования и производные требования» [12] и часть 5 «Направления стандартизации» [13] (остальные части данной серии утверждены в марте-сентябре 2020 г. [14–16]). В феврале 2019 года ИСО/МЭК утверждает основополагающий терминологический стандарт по большим данным. Таким образом, первоначальные международные стандарты были ориентированы, в первую очередь, на гармонизацию разрозненных представлений о больших данных и построение их единой стандартизированной концепции, а также утверждение общих положений, требований к их эталонной архитектуре и обеспечению безопасности данных (см. [17]).

Дальнейшим развитием международной стандартизации больших данных стала разработка в рамках ИСО/МЭК серии стандартов 5259-Х [18–21], посвящённых качеству данных для аналитики и машинного обучения [см. например, 22–23] и описанию концептуальной схемы жизненного цикла работы с данными в системах искусственного интеллекта [24].

Второй этап – стандартизация требований к работе с большими данными в различных сферах деятельности. Этот этап стандартизации в области больших данных связан с разработкой стандартов для отдельных отраслей экономики и сфер деятельности. В первую очередь следует отметить деятельность МСЭ по стандартизации в телекоммуникационной отрасли, связанной с организацией сетей, обеспечивающих передачу больших данных [см. например, 25–26]. Важно отметить, что в процессах принятия стандартизации помимо трех главных международных организаций стандартизации (ИСО, МЭК, МСЭ) активную и не менее значимую роль играют международные организации, занимающиеся в том числе разработкой и утверждением отраслевых стандартов. К их числу можно отнести Организацию по развитию стандартов структурированной информации (Organization for the Advancement of Structured Information Standards, OASIS), Health Level Seven International, Американское общество по испытанию материалов (ASTM International), Консорциум открытых геопространственных данных (Open Geospatial Consortium, OGC) и другие.

Стандартизация больших данных в России началась с существенным отставанием от зарубежных стран и международных организаций, в связи с чем на национальном уровне было принято решение ускоренными темпами сократить разрыв, разработать и ввести в действие ряд основополагающих национальных стандартов в области больших данных, гармонизированных с международными. Для решения этой задачи, равно как и для смежной проблемы развития отечественной стандартизации технологий искусственного интеллекта в 2019 г. был создан технический комитет по стандартизации «Искусственный интеллект» (ТК 164) [27], позиционирующийся как зеркальное отражение на национальном уровне профильного международного подкомитета ISO/IEC JTC 1 SC 42 Artificial Intelligence и объединивший более 100 заинтересованных государственных, коммерческих, научных и образовательных организаций и разработчиков в области технологий искусственного интеллекта и больших данных. Как и на международном уровне, стандартизация больших данных была передана в ведение ТК 164, где изначально была создана рабочая группа 02 «Большие данные», которая в 2020 году – вслед за изменениями в структуре профильного международного подкомитета – была преобразована в подкомитет 02 «Данные» (ТК 164/ПК 02) [28]. Функции секретариата данного подкомитета выполняет Национальный центр цифровой экономики Московского государственного университета имени М.В.Ломоносова. Кроме того, отдельные вопросы стандартизации соответствующих технологий входят в компетенцию технических комитетов по стандартизации «Информационные технологии» (ТК 22), «Криптографическая защита информации» (ТК 26), «Кибер-физические системы» (ТК 194), «Каталогизация продукции» (ТК 430).

Становление парадигмы данных как основы развития цифровой экономики привело к необходимости стандартизации технологий работы с большими данными в рамках развития передовых производственных технологий. Разработка и «локализация» документов по стандартизации больших данных в Российской Федерации были обозначены в качестве задач Национальной технологической инициативы и вошли в План мероприятий («дорожную карту») по направлению «Технет» (передовые производственные технологии) 2018 г. [29]. Позднее для реализации соответствующих мероприятий Министерством промышленности и торговли РФ и Федеральным агентством по техническому регулированию и метрологии был утвержден Перспективный план стандартизации в области передовых производственных технологий на 2018–2025 годы [30], а первоначальный перечень соответствующих стандартов по большим данным был включён в Программу национальной стандартизации на 2019 г. [31].

Впоследствии по причинам низкой экспертной оценки эффективности реализации соответствующей дорожной карты стандартизации «Технета» [32] и необходимости обеспечения единства разрабатываемых стандартов со стандартами в области искусственного интеллекта было принято решение объединить разработку и внедрение стандартов для искусственного интеллекта и больших данных в единой Дорожной карте развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект» [33]. Предполагалось, что реализация данной дорожной карты будет осуществляться в виде серии мероприятий федерального проекта «Нормативное регулирование цифровой среды» [34] в составе национальной программы «Цифровая экономика Российской Федерации», однако ни одно из мероприятий по стандартизации в рамках данного проекта так и не было реализовано. Несмотря на это, следует отметить, что работы по стандартизации больших данных ведутся с 2019 г. в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» по программе Центра

компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных» при Московском государственном университете имени М.В.Ломоносова.

Новый импульс отечественная стандартизация больших данных получила в 2020 г. после принятия Национальной стратегии развития искусственного интеллекта на период до 2030 года [35] и утверждения нового федерального проекта «Искусственный интеллект» [36]. Министерством экономического развития Российской Федерации и Федеральным агентством по техническому регулированию и метрологии была разработана и утверждена Программа стандартизации по приоритетному направлению «Искусственный интеллект» на период 2021–2024 годы [37], которая финансируется из федерального проекта. Мероприятия по стандартизации больших данных являются частью этой программы и направлены на сокращение отставания национальной системы стандартизации от международной.

В настоящее время в России утверждён и вступил в силу гармонизированный ГОСТ Р ИСО/МЭК 20546–2021 «Информационные технологии. Большие данные. Обзор и словарь», положения которого идентичны международному стандарту ISO/IEC 20546:2019 [1]. Кроме того, на различных стадиях разработки находятся проекты национальных стандартов (см. Таблицу 1). Разработка национальных стандартов в сфере работы с большими данными, за исключением одного, закреплена за ТК 164 «Искусственный интеллект».

Таблица 1 – Состояние разработки проектов национальных стандартов Российской Федерации в области больших данных

Наименование проекта национального стандарта	Международный аналог (IDT – идентичный стандарт; MOD – модифицированный стандарт)	Статус разработки
ГОСТ Р «Информационные технологии. Эталонная архитектура больших данных. Часть 1. Структура и процесс применения» [38]	ISO/IEC TR 20547-1:2020 «Information technology – Big data reference architecture – Part 1: Framework and application process» (IDT) [14]	Доработка окончательного варианта и вынесение на голосование в ТК 164 «Искусственный интеллект»
ГОСТ-Р «Информационные технологии. Эталонная архитектура больших данных. Часть 2. Варианты использования и производные требования» [39]	ISO/IEC TR 20547-2:2018 «Information technology – Big data reference architecture – Part 2: Use cases and derived requirements» (IDT) [12]	Утверждение национального стандарта в Росстандарте
ПНСТ «Информационные технологии. Большие данные. Типовая архитектура» ¹ [40]	ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture (MOD) [15]	Утверждение предварительного национального стандарта в Росстандарте
ГОСТ Р «Информационные технологии. Эталонная архитектура больших данных. Часть 5. Направления стандартизации» [41]	ISO IEC TR 20547-5:2018 «Information technology - Big data reference architecture - Part: 5: Standards roadmap» (IDT) [13]	Доработка окончательного варианта и вынесение на экспертизу в ТК 164 «Искусственный интеллект»
ГОСТ Р «Информационные технологии. Большие данные. Техническое задание. Требования к содержанию и оформлению» [42]	Разработан впервые, не имеет международных аналогов.	Утверждение национального стандарта в Росстандарте

¹ ПНСТ «Информационные технологии. Большие данные. Типовая архитектура» закреплён за сферой компетенций ТК 194 «Киберфизические системы».

Наименование проекта национального стандарта	Международный аналог (IDT – идентичный стандарт; MOD – модифицированный стандарт)	Статус разработки
ГОСТ Р «Информационные технологии – Искусственный интеллект – Структура управления процессами аналитики больших данных» [43]	ISO/IEC 24668 Information technology – Artificial intelligence – Process management framework for big data analytics (IDT) [44]	Доработка окончательного варианта и вынесение на голосование в ТК 164 «Искусственный интеллект»
ГОСТ Р «Информационные технологии. Искусственный интеллект. Структура жизненного цикла работы с данными»	ISO/IEC WD 8183 Information technology – Artificial intelligence – Data life cycle framework (MOD)	Сформировано предложение по включению разработки национального стандарта в Программу национальной стандартизации на 2022 г. последующие годы с плановым сроком утверждения в июне 2024 г.
ГОСТ Р «Информационные технологии. Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 1. Обзор, термины и примеры»	ISO/IEC WD 5259-1 Data quality for analytics and ML – Part 1: Overview, terminology, and examples (MOD) [18]	Включён в Программу национальной стандартизации на 2021 г. и последующие годы с плановым сроком утверждения в декабре 2023 г.
ГОСТ Р «Информационные технологии. Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Меры качества данных»	ISO/IEC AWI 5259-2 Data quality for analytics and ML – Part 2: Data quality measures (MOD) [19]	Включён в Программу национальной стандартизации на 2021 г. и последующие годы с плановым сроком утверждения в декабре 2023 г.
ГОСТ Р «Информационные технологии. Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 3. Требования и руководство по управлению качеством данных»	ISO/IEC WD 5259-3 Data quality for analytics and ML – Part 3: Data quality management requirements and guidelines (MOD) [20]	Включён в Программу национальной стандартизации на 2021 г. и последующие годы с плановым сроком утверждения в декабре 2023 г.
ГОСТ Р «Информационные технологии. Качество данных для аналитики и машинного обучения. Часть 4. Структура процесса повышения качества данных»	ISO/IEC WD 5259-4 Data quality for analytics and ML – Part 4: Data quality process framework (MOD) [21]	Включён в Программу национальной стандартизации на 2021 г. и последующие годы с плановым сроком утверждения в декабре 2023 г.

2 Анализ публикационной активности в области стандартизации больших данных

Для достижения поставленной цели исследования, помимо анализа международных и национальных инициатив по стандартизации больших данных был проведен анализ научных публикаций и аналитических материалов по данной теме. В качестве источника научных публикаций рассматривались наиболее авторитетные международные библиографические базы данных Web of Science (WoS) и Scopus, для публикации в которых внимательно отбираются рецензируемые научные издания. Первая отмечается как содержащая тщательно отобранные более качественные источники, вторая – как более широкая по охвату публикаций [45].

На основе консультаций с экспертами в области стандартизации и технологий работы с большими данными были сформированы поисковые образы, учитывающие различные варианты написания терминов. Для наиболее полного отражения картины публикационной активности в сфере стандартизации больших данных был выбран период с 2011 г., когда тематика

стандартизации больших данных стала актуальной, по 2020 г.². При анализе публикационной активности наиболее распространённым и эффективным вариантом проведения поиска является поиск по полям «Тема» (поиск по названию, аннотации, автору и ключевым словам) для WoS³ и «название документа, краткое описание, ключевые слова» (оператор TITLE-ABS-KEY) для Scopus⁴. В соответствии с принятым подходом анализ публикационной активности был осуществлён по следующим этапам.

1. Из «ядерных» коллекций WoS и Scopus был выделен массив публикаций по большим данным, определённый по следующему перечню ключевых слов и словосочетаний: big data; data analytic; data mining; data science; descriptive analytic; diagnostic analytic; Hadoop; large dataset; MapReduce; massive data; predictive analytic; prescriptive analytic; semi-structured data; text mining; unstructured data. Детальное описание процедуры формирования приведенного выше поискового образа описан в статье [46].

За временной период 2011–2020 гг. поисковый запрос выдал 149 349 результатов в базе WoS и 289 514 результатов в базе Scopus.

2. На следующем этапе из сформированного массива публикаций осуществлялся отбор публикаций, имеющих отношение к стандартизации технологий работ с большими данными, путём задания специализированного поискового образа со следующими ключевыми словами и словосочетаниями: committee-based standardization; consortia standard; data quality; data standards; formal standard; government-based standardization; international standard; ISO/IEC; multi-mode standardization; national standard; proprietary standard; reference architecture; standard; standardization union; standardization; standards system.

Отбор указанных ключевых слов и словосочетаний основывался на предварительном исследовании, включающем в себя анализ релевантных публикаций и формирование исходного перечня специфичных для данной области ключевых слов с участием экспертов. Приведённые ключевые слова и словосочетания позволяют получить максимальное покрытие исследуемой области, однако дают значительное количество «шума» – публикаций, слабо либо вообще не связанных с темой исследования, что подтверждается экспериментальной работой по ознакомлению с отобранным массивом публикаций и экспертной оценкой поисковых выданных. Это объясняется, в первую очередь, особенностями и многозначностью использования термина «стандарт» и производных от него в различных отраслях науки. Например, термин «стандарт» и производные от него могут использоваться для обозначения как собственно стандарта как нормативно-технического документа, так и эталонных (стандартизированных) подходов, практик (например, «стандартной» модели качества данных, или «стандартной» практики лечения, или «стандарта» качества услуги, или «золотого стандарта»); общепринятого, обычного подхода, не имеющего формализованной формы (например, «стандартизированные» вопросы; единообразного формата или типа (например, «стандартный» формат файла); этап обработки больших данных (например, «стандартизация» данных как обозначение приведения их к единому формату).

Поисковые запросы первого и второго этапов были объединены логическим оператором «AND», а поиск по ним за временной период 2011–2020 гг. дал следующие результаты: 11 122 публикации (из них 553 обзорные) в базе данных WoS и 21 194 публикации (из них 828 обзорные) в базе данных Scopus. Распределение указанных публикаций по годам в разрезе баз данных представлено на рисунке 1. Интересно при этом отметить, что за временной период с 1998 г. (когда был введён в научный оборот термин «большие данные» (big data)) по 2010 г. было найдено 2 062 публикации (из них 57 обзорные) в WoS и 4 096 публикаций (из них 161 обзорная) в Scopus.

² Данные о публикациях могут подгружаться и обрабатываться в реферативных базах с задержкой, в связи с чем было принято решение не учитывать публикации 2021 г.

³ Показатели публикационной активности для базы данных Web of Science приведены на основе выгрузок, произведённых 20.08.2021.

⁴ Показатели публикационной активности для базы данных Scopus приведены на основе выгрузок, произведённых 27.08.2021.

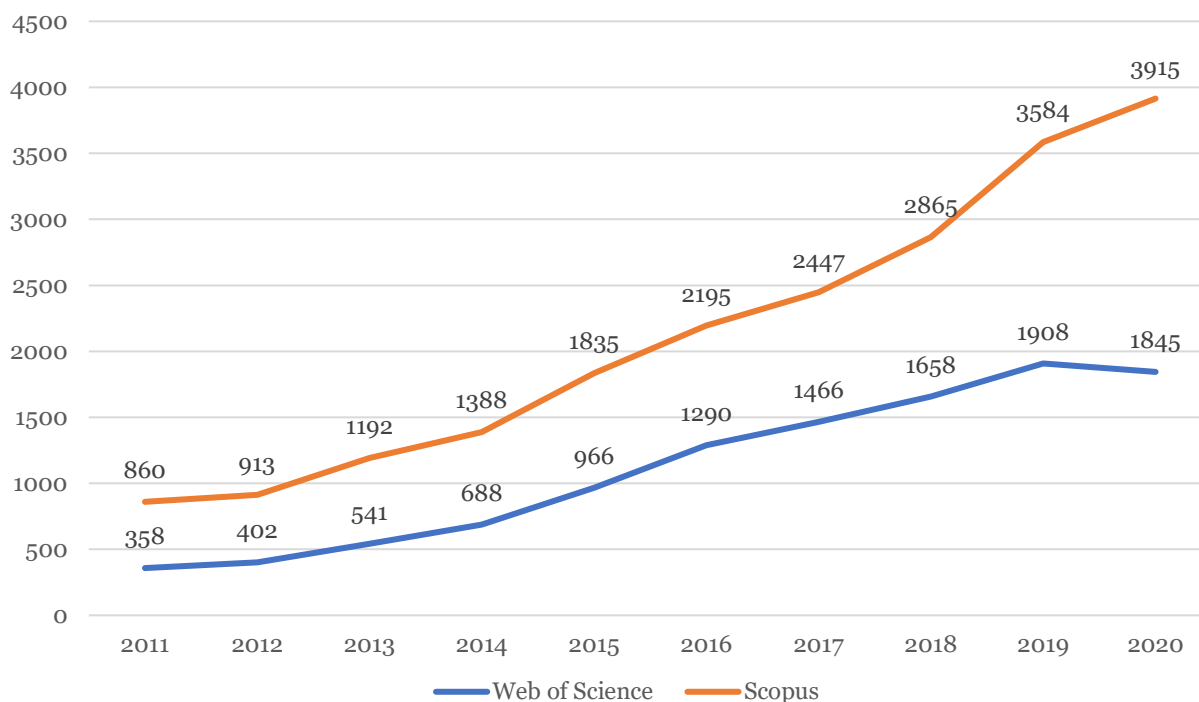


Рисунок 1 – Показатели публикационной активности по стандартизации больших данных, 2011–2020 гг.
Источники: WoS, Scopus

Приведённый график демонстрирует устойчивый рост публикационной активности в сфере стандартизации технологий работы с большими данными; при этом можно отметить незначительное падение в WoS за 2020 г. что, скорее всего, объясняется процедурой индексирования статей, которая может продолжаться в течение полутора-двух лет после публикации.

Выборочный анализ наименований и аннотаций ста научных публикаций из найденного массива показал, что большинство публикаций не относится к предмету настоящего исследования, в связи с чем было принято решение для публикаций, имеющих отношение к стандартизации технологий работ с большими данными провести поиск только по полю «название публикации», что признаётся допустимой поисковой стратегией [47]. В связи с этим было принято решение сузить поисковый запрос по следующим параметрам: для публикаций в сфере больших данных оставить поиск в пределах «темы» в базе WoS или «названия документа, краткого описания, ключевых слов» (оператор TITLE-ABS-KEY) в базе Scopus, для публикаций, относящихся к стандартизации больших данных, применить поиск в пределах параметров «заголовков» и «ключевые слова автора» в базе WoS и «название документа» (оператор TITLE) и ключевые слова (оператор KEY) в базе Scopus. Такой поисковый запрос за временной период 2011–2020 гг. выдал 1140 результатов, из них 322 результата (включая 18 обзорных публикаций) в базе WoS и 818 результатов (включая 41 обзорная публикация) в базе Scopus.

Сравнительный анализ наименований, аннотаций и ключевых слов ста случайно выбранных публикаций, найденных на предыдущем этапе, и публикаций, найденных на данном этапе, показал, что наиболее релевантные результаты даёт последний вариант поискового запроса.

На следующем этапе был проведён анализ полученных результатов поиска на предмет соответствия исследуемой теме. Первоначальное выборочное ознакомление с наименованиями, ключевыми словами, аннотациями и полными текстами показало, что в большинстве публикаций предмет исследования либо значительно отличается, либо в малой степени затрагивает вопросы стандартизации технологий работы с большими данными (как правило, на уровне упоминаний в тексте). Исходя из этого был проведён анализ полной выборки публикаций по наименованиям и аннотациям. В результате было отобрано 305 уникальных публикаций за период 2011–2020 гг.,

относящихся к тематике стандартизации больших данных (см. рисунок 2), которые учитывались при анализе публикационной активности.

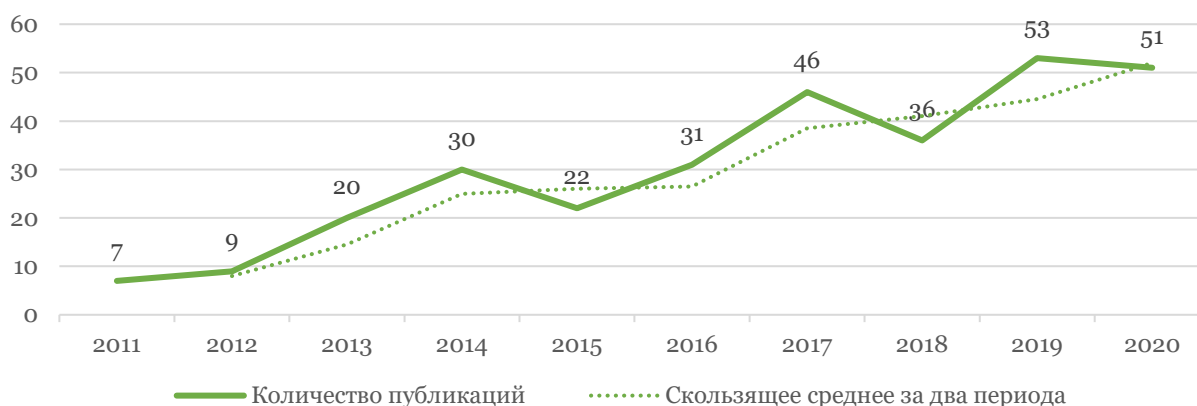


Рисунок 2 – Специализированные публикации по стандартизации больших данных, 2011–2020 гг.
Источники: WoS, Scopus

Общее количество специализированных публикаций имеет устойчивую тенденцию постепенного роста с 2011 г. по 2020 г., за исключением периодов снижения публикационной активности в 2015 г. и 2018 г. Представляется, что отмеченный рост имеет прямую зависимость от истории становления и развития сферы стандартизации больших данных (см. раздел 1). Наиболее заметный рост публикационной активности в 2013–2014 гг. связан с принятием первых стандартов в области больших данных Национальным институтом стандартов и технологий США (NIST) [5–7]. Аналогичный всплеск публикационной активности в 2019–2020 гг. связывается с началом работ по международной стандартизации больших данных МСЭ и Объединенным техническим комитетом №1 ИСО/МЭК.

Анализ тематической направленности публикационного массива показал, что все статьи разбиваются на две категории: общесистемные вопросы стандартизации больших данных (такие как терминологическая база, эталонная архитектура, качество данных, анализ данных, безопасность/конфиденциальность при работе с данными) и стандартизация работы с большими данными в отдельных сферах деятельности (здравоохранение, государственное управление, бизнес, промышленность, наука и другие). Распределение публикаций по тематике приведено ниже на рисунках 3 и 4 соответственно.

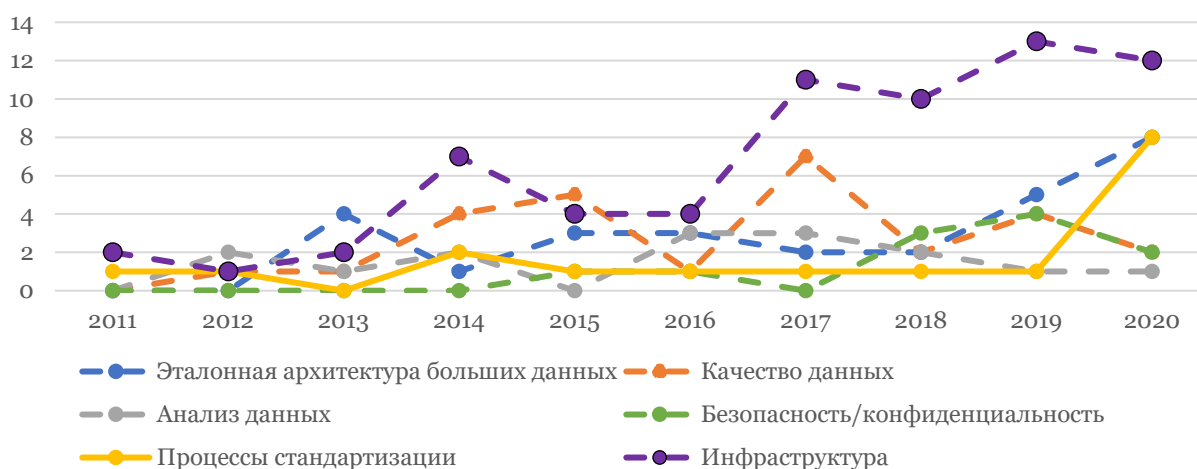


Рисунок 3. Публикационная активность в сфере стандартизации технологий, решений и сервисов работы с большими данными, 2011–2020
Источники: WoS, Scopus

На рисунке 3 представлена динамика публикационной активности, связанной со стандартизацией технологий, решений и сервисов работы с большими данными: инфраструктура работы с большими данными, включая информационные системы (66 статей); эталонная архитектура больших данных (28); качество больших данных (27); аналитика больших данных (15); безопасность/конфиденциальность при работе с большими данными (11); процессы стандартизации в сфере больших данных (17). В категорию «процессы стандартизации» объединены публикации по истории и методологии разработки и ввода в действие стандартов в области больших данных, практике применения стандартов, а также стандартизации процессов взаимодействия и обмена данными.



Рисунок 4 – Публикационная активность в сфере стандартизации работы с большими данными в отдельных сферах деятельности, 2011–2020

Источники: WoS, Scopus

На рисунке 4 представлена динамика публикационной активности, связанной со стандартизацией работы с большими данными в четырех сферах деятельности с максимальным числом публикаций. В целом следует отметить тенденцию увеличения числа таких публикаций, что отражает повышение спроса на стандартизацию больших данных в различных прикладных экономических и социальных аспектах. Анализ динамики демонстрирует стабильно высокий интерес экспертов в сфере стандартизации технологий работы с большими данными в здравоохранении, включая вопросы создания и использования медицинских баз данных. При этом постепенно увеличивается, хоть и менее интенсивно, количество публикаций, связанных со стандартизацией использования больших данных в системе государственного управления и в науке.

Подводя итоги, можно сделать несколько выводов о тенденциях публикационной активности в сфере стандартизации больших данных:

- на период 2015–2016 гг. приходится первый пик публикационной активности по стандартам в области работы с большими данными, что связано, в первую очередь, с появлением доклада «Большие данные» [9] и началом разработки международных стандартов; второй пик приходится на 2019–2020 гг. в связи с принятием терминологического стандарта ИСО/МЭК 20546 [1] и серии стандартов ИСО/МЭК 20547-X «Эталонная архитектура больших данных» [12–16];
- в последние пять лет происходит «расщепление» публикационной активности на тематику, связанную с общесистемными вопросами стандартизации больших данных, сопровождающуюся уменьшением числа таких публикаций и увеличением числа публикаций о стандартизации работы с большими данными в отдельных отраслях экономики, сферах социальной жизни и системы государственного управления (исключением сферы здравоохранения, где публикационная активность остается стабильно высокой);
- начиная с 2014 г. возрастает количество публикаций в области стандартов по различным аспектам работы с данными и, прежде всего, стандартов на качество данных, в том числе в различных отраслях (см., например, [48–51]);

- начиная с 2017 г. начинается активная подготовка и публикация результатов исследований по построению эталонных архитектур больших данных для различных сфер деятельности (см. например, [52–54]).

В настоящее время в России ведется активная работа по разработке и гармонизации общесистемных стандартов в области работы с большими данными, поэтому в рамках данной статьи вопросы стандартизации работы с большими данными в конкретных сферах деятельности не рассматриваются. Дальнейшее исследование будет использовать результаты проведенного анализа публикационной активности, действующие международные стандарты и проекты стандартов (прежде всего – ИСО/МЭК), а также проекты национальных стандартов Российской Федерации в области больших данных.

3 Большие данные: терминология и направления стандартизации

Зонтичный характер базового понятия «большие данные» породил две серьезные проблемы для потенциальных разработчиков технологий для работы с большими данными. Первая проблема состоит в многозначности и неопределенности самого термина, о которых шла речь во введении к данной статье, а вторая проблема связана с необходимостью единого подхода к стандартизации архитектуры системы для работы с большими данными.

Именно этими проблемами в первую очередь озаботилась рабочая группа «Большие данные» (впоследствии переименованная в рабочую группу «Данные») подкомитета 42 «Искусственный интеллект» Объединенного технического комитета №1 ИСО/МЭК. В 2019 г. была завершена трехлетняя работа над международным терминологическим стандартом «Большие данные. Обзор и словарь» [1], а в 2020 г. – работа над серией стандартов «Эталонная архитектура больших данных» [12–16], которые подробно рассматриваются далее.

Терминологический стандарт ISO/IEC 20546:2019 «Information technology – Big data – Overview and vocabulary» [1] представляет собой компромисс между несколькими конкурирующими подходами к определению больших данных и связанными с ними технологиями и системами. За основу было взято определение, предложенное в пионерной работе аналитика Meta Group Дага Лэйни [55] для характеристики больших данных через объём, скорость изменений и разнообразие (3V = Volume, Velocity, Variety). Позднее делались попытки добавить еще несколько V (например, V = Volatility, V = Veracity, V = Validity) к характеристикам больших данных – в качестве курьеза можно отметить, что в отдельных статьях насчитывалось более сорока подобных V. Однако в международный стандарт в результате консенсуса были включены только три основные характеристики, где разнообразие могло меняться на вариативность (V = Variability), именно это определение приведено во Введении. В нем и далее определения терминов на русском языке взяты из идентичного национального стандарта «Информационные технологии. Большие данные. Обзор и словарь» [2], вступившего в действие с 1 ноября 2021 года.

Отметим, что наряду с основными характеристиками:

- объем данных (data volume) – количественная характеристика данных (3.1.5), влияющая на выбор ресурсов для вычислений и хранения, а также на управление данными в процессе обработки;
- скорость обработки данных (data velocity) – скорость потока, с которой данные создаются, передаются, сохраняются, анализируются или визуализируются;
- разнообразие данных (data variety) – диапазон форматов, логических моделей, временных шкал и семантики массива данных;
- вариативность данных (data variability) – изменения в скорости передачи, формате или структуре, семантике или качестве массива данных;

в стандарте определены и другие встречающиеся характеристики, такие как:

- достоверность данных (data veracity) – полнота и/или точность данных;
- изменчивость данных (data volatility) – характеристика данных, относящаяся к скорости их изменения с течением времени.

Наряду с характеристиками больших данных в стандарте ISO/IEC 20546:2019 даны определения основных понятий, относящихся к технологиям работы с большими данными, таких как распределенная обработка данных, аналитика данных, облачные вычисления, горизонтальное и вертикальное масштабирование и т.д.

Заканчивается международный стандарт кратким описанием ключевых процессов обработки данных.

С терминологическим стандартом ISO/IEC 20546 тесно связан международный технический отчёт ISO/IEC TR 20547-5: 2018 «Information technology» – Big data reference architecture – Part 5: Standards roadmap» [13] который, прежде всего, содержит сводку основных стандартов, связанных с эталонной архитектурой больших данных.

В отчете также приводится достаточно представительный (хотя и не исчерпывающий) перечень официальных организаций стандартизации, отраслевых консорциумов и организаций-разработчиков программного обеспечения, которые заинтересованы в стандартизации больших данных на международном и национальном уровнях.

Последние разделы отчета посвящены потенциальным пробелам в стандартизации больших данных в таких областях как сценарии использования больших данных; спецификации и стандартизация метаданных, включая происхождение данных; методы обработки (например, пакетные, потоковые); семантика соответствия больших данных предметным областям; расширенные сетевые протоколы для эффективной передачи данных; общие и предметные онтологии и таксономии для описания семантики данных, включая взаимосвязь между онтологиями;

Стандарты обычно создаются с учетом лучших практик и подходов, которые проверены на реальных приложениях и в теории. В случае больших данных многие стандарты также развиваются на основе существующих стандартов, которые модифицируются для учета уникальных особенностей больших данных. Это направление деятельности обозначено в техническом отчете как подход, который принят и реализуется на международном уровне.

4 Большие данные: структура и процесс применения

Как уже отмечалось выше, еще одним барьером к развитию и использованию технологий работы с большими данными было отсутствие референсной архитектуры информационных систем для оперирования большими данными. В этом направлении в ИСО/МЭК с 2018 года началась активная разработка единой серии стандартов ISO/IEC 20547-X [12–16], посвящённых эталонной архитектуре больших данных, которая была успешно завершена в 2020 г.

Стандарты серии ISO/IEC 20547-X предназначены в первую очередь для обеспечения однозначного описания архитектуры системы работы с большими данными. Состав и взаимосвязи между отдельными частями серии стандартов представлен в ISO/IEC 20547-1 [14] (см. рисунок 5).

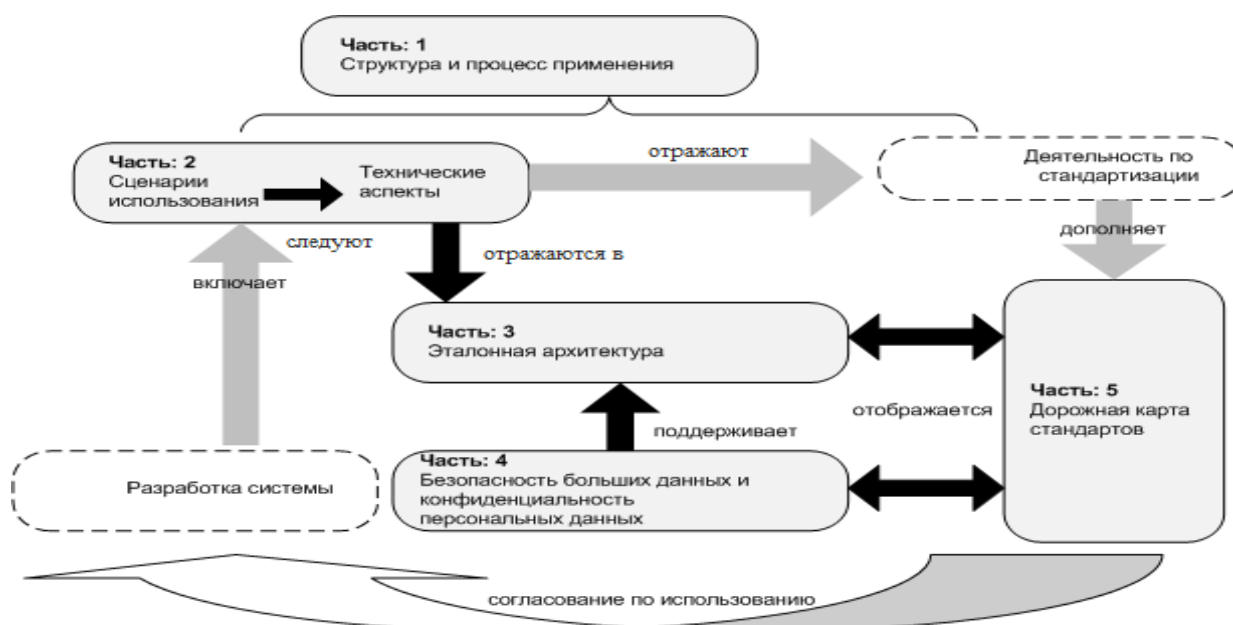


Рисунок 5 – Взаимосвязи между частями стандарта ИСО/МЭК 20547-X
Источник: [14]

В *части 1* «Структура и процесс применения» представлена логическая взаимосвязь между всеми частями стандарта ISO/IEC 20547-X, дано описание (концептуальной схемы) эталонной архитектуры больших данных и процесса применения стандарта в конкретной предметной области. В *части 2* «Сценарии использования и производные требования» представлены примеры описания сценариев (вариантов) использования больших данных в различных предметных областях, а также выделены вытекающие из них технические требования к системам для работы с большими данными. В *части 3* «Эталонная архитектура» содержится описание эталонной архитектуры больших данных, включающей в себя основные архитектурные понятия, в том числе пользовательское и функциональное представление. В *части 4* «Безопасность больших данных и конфиденциальность персональных данных» представлено описание аспектов доверия и безопасности при работе с большими данными, а также обеспечения конфиденциальности персональных данных применительно к эталонной архитектуре больших данных. Обоснованы причины снижения уровня безопасности и конфиденциальности при работе с большими данными, а также представлены подходы, направленные на повышение уровня информационной безопасности систем больших данных. С учетом особенностей работы с большими данными и их высокой ценности сформулированы задачи обеспечения безопасности и конфиденциальности персональных данных, а также представлены рекомендации по обеспечению безопасности и защите персональных данных при работе с большими данными как на уровне организации, так и на уровне экосистемы больших данных. В *части 5* «Дорожная карта стандартов» содержится перечень стандартов (как существующих, так и разрабатываемых), организаций, активно участвующих в процессах стандартизации, а также представлен анализ существующих проблем и направлений разработки будущих стандартов, относящихся к большим данным.

Общее описание системы для работы с большими данными, содержащееся в ISO/IEC 20547-1 позволяет сформулировать ключевые требования к их функционированию:

- обеспечение эффективной обработки, хранения, управления и анализа больших массивов данных, характеризующихся объёмом, разнообразием, скоростью обработки, а также их вариативностью; с этой целью должны использоваться различные технологии масштабирования;
- реализация перспективных методик построения масштабируемых систем данных на основе независимых ресурсов в ситуациях, когда характеристики массивов данных требуют разработки специальных архитектур для эффективного хранения, обработки и анализа;
- реализация парадигмы распределения массивов данных по горизонтально связанным и независимым ресурсам с целью достижения масштабируемости, необходимой для эффективной обработки больших массивов данных.

Разнообразие систем больших данных и технологий работы с ними определяют набор необходимых и достаточных требований, предъявляемых к эталонной архитектуре, позволяющих реализовывать широкий спектр потенциальных сценариев использования больших данных.

Вторая половина стандарта ISO/IEC 20547-1 посвящена описанию пошагового процесса применения эталонной архитектуры для разработки архитектуры конкретной системы больших данных. Эталонная архитектура больших данных, описанная в стандарте ISO/IEC 20547-3 (см. далее раздел 6), является достаточно общей и предназначена для различных сценариев использования, однако для учета потенциального разнообразия систем больших данных (и их компонентов) процесс применения предоставляет возможность расширения эталонной архитектуры. Основной особенностью этого расширения является идентификация дополнительных действий, связанных с ролями, и/или назначение действий различным ролям/подролям.

Применение эталонной архитектуры к построению конкретной системы больших данных сводится к выполнению ряда взаимосвязанных шагов.

1. Идентификация заинтересованных сторон и их требований.
2. Отображение в ролях и подролях заинтересованных сторон и их требований.
3. Разработка подробных описаний деятельности и её соответствие интересам (требованиям).
4. Определение функциональных компонентов для реализации деятельности.
5. Определение соответствия сквозных действий/функциональных компонентов интересам (требованиям).

Каждый из перечисленных шагов достаточно детально описывается в стандарте ISO/IEC 20547-1 с отсылкой к другим действующим стандартам.

На первом шаге в процессе применения эталонной архитектуры осуществляется выявление заинтересованных сторон и определение их интересов, связанных с разработкой системы больших данных.

На следующем шаге решается задача отображения заинтересованных сторон и их требований в общей структуре понятий и представлений о системе больших данных в ролях и подролях, под которыми понимаются виды деятельности с большими данными.

На третьем шаге выполняется подробное описание деятельности для конкретной системы больших данных, представленной в категориях ролей и подролей с учетом требований заинтересованных сторон.

На следующем шаге, представляющем этап высокоуровневого проектирования системы больших данных, выполняется идентификация функциональных компонентов, предназначенных для осуществления деятельности системы больших данных.

На последнем шаге процесса разработки выполняется валидация соответствия сформированных функциональных компонентов системы больших данных заданным требованиям путём трассировки каждого требования до функционального компонента и наоборот, трассировки каждого функционального компонента до конкретного требования в рамках реализуемой деятельности.

5 Сценарии использования больших данных

Варианты использования технологий работы с большими данными, а также вытекающие из них требования к эталонной архитектуре создаваемых систем, зафиксированы в международном техническом отчёте ISO/IEC 20547-2 [12], который содержит описания 51 варианта использования больших данных, классифицированных по следующим 9 категориям:

- деятельность государственных органов;
- коммерческая деятельность;
- оборона;
- здравоохранение и медико-биологические науки;
- глубокое обучение (Deep Learning) и социальные сети;
- экосистема для исследований;
- астрономия и физика;
- науки о Земле, экологические науки и полярные исследования;
- энергетика.

Появление любой инновационной технологии влечёт за собой желание собирать и распространять интересные варианты её использования. Этим занимаются многочисленные профессиональные ассоциации, научные учреждения и консультационные фирмы, а в последнее время – международные организации по стандартизации. Усилия по сбору, описанию и анализу вариантов использования можно разделить на две части: 1) «коллекционирование» кейсов и 2) попытки проанализировать собранные варианты и обнаружить какие-либо высокоуровневые закономерности. При этом социально-экономические эффекты от «коллекционирования» кейсов проявляются лишь в первые несколько лет с момента появления технологии в пилотных проектах, но в дальнейшем сходят на нет. Высокоуровневый анализ кейсов редок ввиду нехватки сильных аналитиков.

К сожалению, все проблемы, связанные с «коллекционированием» вариантов использования, можно наблюдать и в техническом отчёте ISO/IEC 20547-2. В этом документе недостаёт высокоуровневой аналитики, доведённой до чётко изложенных результатов, пригодных для использования лицами, принимающими решения. В этом есть потенциал для разработки в будущем небольшого по объёму отечественного стандарта, закрывающего данный пробел.

Вместе с тем адаптация технического отчёта ISO/IEC 20547-2 принесёт пользу отечественным специалистам и организациям, ввиду следующего:

- российские специалисты смогут на родном языке познакомиться с материалами, отражающими внедрение и развитие технологий работы с большими данными за рубежом;

- описания многих вариантов использования представляют самостоятельный интерес, особенно для специалистов из «родственных» сфер деятельности – здесь можно почерпнуть полезные идеи для собственных проектов;
- в документе сделана попытка выделить типичные проблемы, пути их решения и варианты дальнейшего развития технологий работы с большими данными, которые могут стать полезной отправной точкой для последующей аналитики накопленного опыта;
- в документе нашли отражение вопросы обеспечения сохранности, доступности для конечных пользователей и эффективного повторного использования разнородных данных, накапливаемых в рамках научных исследований и экспериментов – то есть, по сути, речь часто идёт о стратегическом управлении информацией и данными с целью оптимизации расходов и увеличения полезной отдачи;
- документ отражает хороший мировой опыт подготовки и обработки описаний вариантов использования – так, ИСО на основе данного документа подготовила технический отчёт о вариантах использования технологий распределённых реестров (блокчейна), известно об интересе к этому опыту и других международных технических комитетов.

Если говорить о связанных с большими данными проблемах и рисках, то из описаний вариантов использования можно увидеть, в том числе, следующие.

«Хрупкость» результатов – даже небольшие изменения в составе данных и в алгоритмах обработки способны привести кардинальным отличиям в результатах. Это особенно критично при использовании больших данных для принятия юридически значимых решений, поскольку в случае спора часто невозможно эти результаты воспроизвести.

Риски для информационной безопасности и персональных данных. В последнее время ужесточаются требования к информационной безопасности и особенно к защите персональных данных. При этом именно технологии работы с большими данными и искусственного интеллекта создают новые риски, заставляя ограничивать доступ к разнородным массивам данных и их объединение или совместную обработку, а также проводить анонимизацию. Всегда следует помнить, что в сфере больших данных значительное конкурентное преимущество (в том числе при обеспечении национальной безопасности) получает сторона, обладающая наиболее мощной инфраструктурой для обработки данных, ресурсами, кадрами и т.д.

Проблемы документирования и долговременной сохранности. Для решений на основе больших данных существует проблема объяснимости принятых решений и их полноценного документирования, а также обеспечения целостности, аутентичности, надёжности, конфиденциальности и пригодности к использованию ценных данных в долговременной перспективе. Многие проблемы связаны с быстрым устареванием программной, аппаратной и сетевой инфраструктуры для работы с данными, а также с необходимостью постоянно инвестировать существенные ресурсы в её поддержание и развитие.

Правовые риски. Постоянно появляются противоречия между требованиями к эффективности архитектуры больших данных и требованиями законодательства соответствующих юрисдикций. Глобальное распределённое хранение больших данных может, например, противоречить требованиям законодательства к локализации данных, особенно персональных. Проблематичным может оказаться исполнение повышенных требований к импортозамещению.

Проблемы интерпретации. Данные могут быть правильно интерпретированы только в правильном контексте. При объединении разнородных наборов данных нередко имеет место частичная или даже полная потеря ключевого контекста (например, сведений о том, кто, когда, каким образом и для чего собрал данные, какова их точность, насколько они актуальны и т.д.). В результате даже сами по себе вполне корректные данные могут быть неверно интерпретированы (в том числе умышленно), что может приводить к принятию ошибочных деловых и политических решений.

Проблемы контролируемого сбора данных. Общий принцип обработки данных заключается в том, что лишние данные мешают эффективной работе алгоритмов и «поедают» ресурсы. Во многих проектах большое внимание уделяется оперативной предобработке поступающих данных, результаты которой используются для оперативного управления процессом сбора данных.

В техническом отчёте ISO/IEC 20547–2 также имеется раздел с перечислением технических проблем, выявленных по результатам анализа вариантов использования. Таким образом, выборка

и анализ собранных вариантов использования позволили – через производные требования из них – построить эталонную архитектуру больших данных (см ниже раздел 6). Обобщённые требования и анализ вариантов использования также позволили разработать проект национального стандарта о требованиях заказчика к действиям, связанным с использованием больших данных (см. ниже раздел 8), в котором отражено, что соответствующие требования укладываются в производные требования, зафиксированные стандартом ISO/IEC 20547-2, и гарантируют качественный результат.

6 Большие данные: эталонная архитектура

Эталонные архитектуры разрабатываются для решения широкого круга задач, и их основное предназначение состоит в ориентации на будущее и использование в качестве основы для будущих реализаций информационных систем [56]. Архитектура систем для работы с большими данными позволяет принимать, обрабатывать и анализировать данные, которые являются слишком объёмными или слишком сложными для традиционных информационных систем, таких как реляционные базы данных. Эталонная архитектура предоставляет заинтересованным сторонам универсальный язык для описания больших данных, обеспечивает поддержку общих стандартов, спецификаций и шаблонов, а также демонстрирует последовательность действий при реализации технологий для решения однотипных задач.

Результаты применения эталонной архитектуры позволяют снять или существенно уменьшить возникающие проблемы при эксплуатации систем для работы с большими данными и дают инструмент для описания, обсуждения и развития специализированных архитектур. В данном разделе рассматривается эталонная архитектура больших данных, описанная в третьей части международного стандарта ISO/IEC 20547-3 [15].

Архитектура больших данных обычно используется для реализации следующих сценариев [57]:

- хранение и обработка данных, в том числе неструктурированных, в объёмах, слишком больших для традиционной базы данных;
- преобразование неструктурированных или слабо структурированных данных для анализа и создания отчетов;
- запись, обработка и анализ непривязанных потоков данных в режиме реального времени или с низкой задержкой;
- прогнозная аналитика и машинное обучение.

Обычно разработчики решений для работы с большими данными стремятся объединить возможности обработки разнородных данных. Одним из распространённых подходов является использование платформы Hadoop, которая работает по принципу перемещения вычислений ближе к месту хранения данных: обработка обычно выполняется на больших кластерах серверов, созданных с помощью стандартного аппаратного обеспечения [58]. Сочетание платформы Hadoop со стандартными серверами – основа для экономичной и высокопроизводительной аналитической платформы для параллельной работы приложений. Среди сформировавшихся подходов к работе с большим данными можно отметить модель распределённых параллельных вычислений MapReduce, которая находит применение в компьютерных кластерах [59]. Согласно этой модели, приложение разделяется на большое количество одинаковых элементарных заданий, выполняемых на узлах кластера и затем естественным образом формирующих конечный результат. Для формирования сложных и гибко построенных запросов к большим данным используются языки NoSQL (от англ. Not Only SQL, не только SQL). Обычные реляционные базы данных обеспечивают формирование для достаточно быстрых и однотипных запросов. В случае больших данных нагрузка может существенно превышать разумные пределы и использование СУБД становится неэффективным.

Разнообразная природа вариантов использования больших данных (см. выше раздел 5) определяет необходимость того, что эталонная архитектура больших данных должна быть достаточно общей и охватывать многообразие потенциальных архитектур систем работы с большими данными. С объектно-ориентированной точки зрения она может быть представлена как абстрактный класс, определяющий структуру и атрибуты конкретных вариантов архитектур. Таким образом, эталонная архитектура должна включать структуру и взаимосвязь компонентов, правила и ограничения, общие для всех систем больших данных, а также ряд соглашений,

принципов и практик для описания архитектур систем больших данных, в том числе отражать взаимосвязи с окружающей средой, областью применения, заинтересованными сторонами и др.

На рисунке 6 представлена концептуальная схема взаимосвязей между базовыми понятиями эталонной архитектуры больших данных. Эталонная архитектура формируется для конкретной области применения, связанной с большими данными и определяющей окружающую среду. В случае больших данных окружающая среда описывается и определяется ключевыми характеристиками больших данных: объемом, скоростью обработки, разнообразием, а также их вариативностью. Окружающая среда включает заинтересованные стороны и их интересы. Под заинтересованными сторонами подразумеваются пользователи, владельцы, архитекторы и другие субъекты любой системы, у которых имеется интерес, связанный с большими данными.

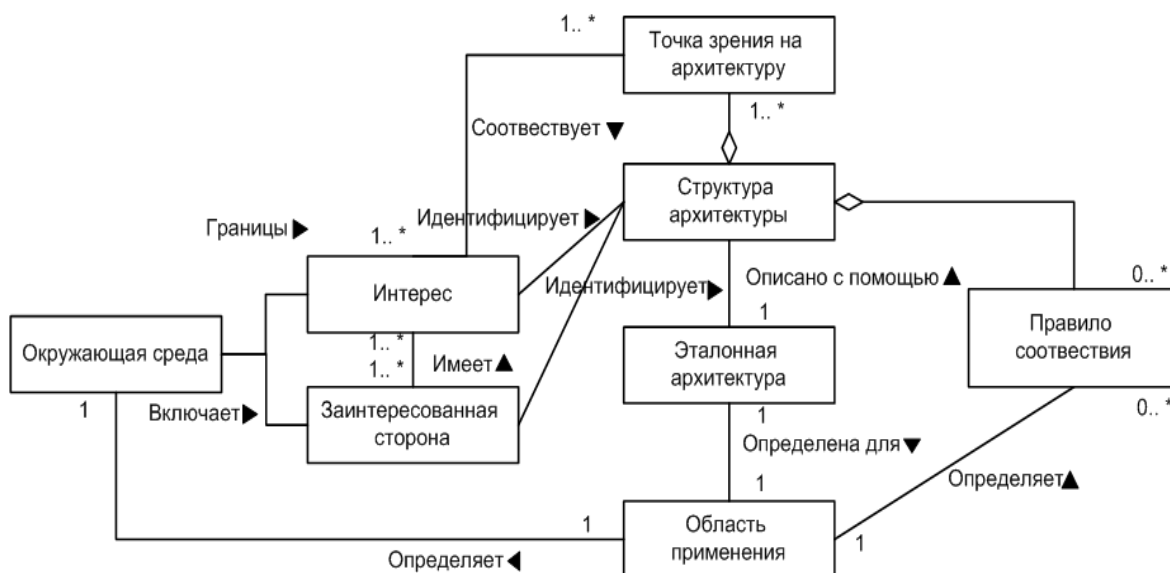


Рисунок 6 – Концептуальная схема взаимосвязей между базовыми понятиями эталонной архитектуры
Источник: [14]

Эталонной архитектуре соответствует концептуальная модель экосистемы больших данных, определяющая роли/подроли и их отношения в экосистеме, а также описание типов деятельности ролей и подролей в экосистеме больших данных.

В основе архитектуры систем больших данных лежат логические отношения между такими сущностями, как: стороны, сквозные аспекты, роли/подроли, деятельности и функциональные компоненты, которые составляют архитектуру системы больших данных (рисунок 7).

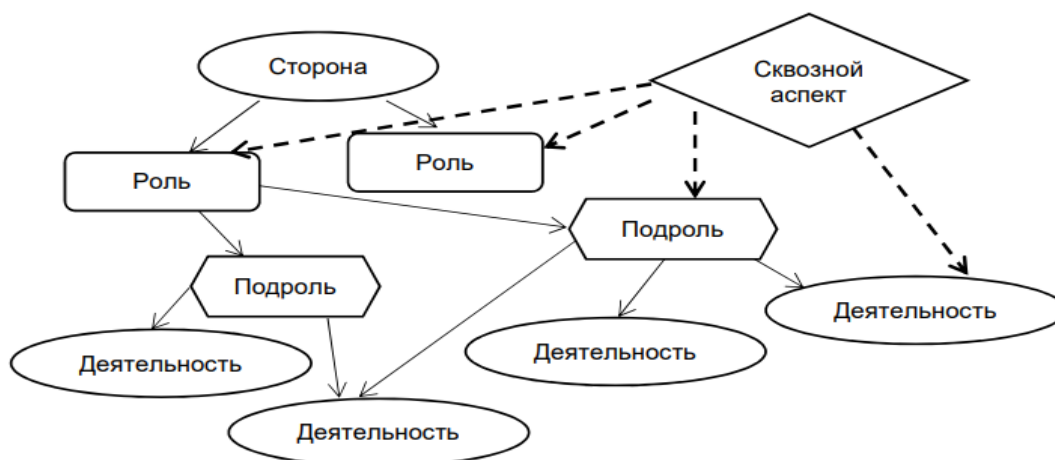


Рисунок 7 – Логические отношения между сущностями архитектуры больших данных в представлении пользователя. Источник: [15]

Сторона представляет собой физическое или юридическое лицо, или группу лиц, которые в экосистеме больших данных представляются её заинтересованными сторонами.

Под ролью понимается набор деятельностей с большими данными. Подмножество деятельностей с большими данными для конкретной роли носят название подроли, при этом деятельности данной роли могут совместно использовать различные подроли.

Деятельность представляет собой определенное исполнение одной задачи или набора задач. Деятельности с большими данными должны иметь цели и обеспечивать получение одного или нескольких результатов, которые достигаются с использованием функциональных компонентов.

В экосистеме больших данных реализуются так называемые сквозные аспекты, которые обеспечивают возможность координации между ролями и функциональными компонентами. Они влияют на выполнение нескольких ролей, действия с большими данными и функциональные компоненты, а также учитываются при совместном выполнении конкретных ролей или использовании функциональных компонентов.

Концепция эталонной архитектуры больших данных, представленная в стандарте ISO/IEC 20547–3, является основой для общего описания систем работы с большими данными и рассматривается с двух точек зрения:

- *пользовательского представления*, включающего роли, подроли, деятельности и сквозные аспекты функционирования системы работы с большими данными, обеспечивающие удовлетворение потребностей заинтересованных сторон;
- *функционального представления*, включающего функциональные уровни, компоненты и многоуровневые функции, обеспечивающие реализацию действий и сквозных аспектов, указанных в представлении пользователя.

Каждая из точек зрения, в свою очередь, затрагивает один или несколько интересов (проблем, требований). В рамках указанных архитектурных представлений интересы могут быть воплощены в одной или нескольких ролях, деятельностях и функциональных компонентах. Схема взаимодействия заинтересованных сторон, а также интересы (проблемы, требования), связанные с указанными представлениями, представлены на рисунке 8.

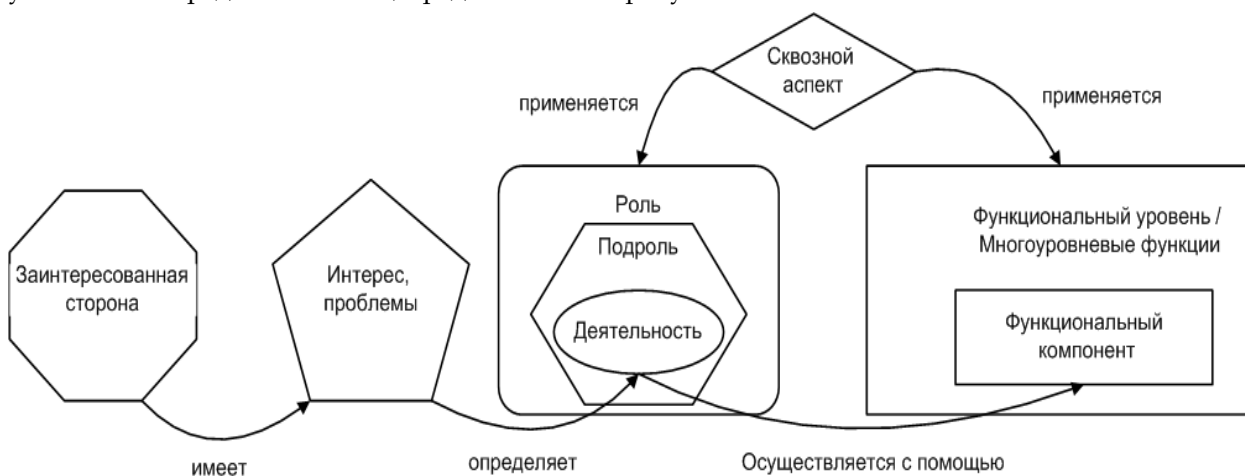


Рисунок 8 – Схема взаимодействия между компонентами представлений эталонной архитектуры больших данных
Источник: [15]

Сквозные аспекты, координирующие совместное выполнение отдельных ролей или использование функциональных компонентов, существуют как в пользовательском, так и функциональном представлениях эталонной архитектуры больших данных. Примером сквозного аспекта может служить обеспечение безопасности и конфиденциальности персональных данных. В отличие от пользовательского функциональное представление является технологически нейтральным представлением о функциях, необходимых для формирования системы работы с большими данными.

Функциональное представление описывает распределение функций, обеспечивающих поддержание деятельности с большими данными. Зависимости между функциями определяет функциональная архитектура. Функциональное представление включает следующие понятия в архитектуре больших данных (рисунок 9):

- **функциональные компоненты**, которые являются функциональными строительными блоками, необходимыми для участия в деятельности;
- **интерфейсы функциональных компонентов**, выполняющие функции границы между двумя функциональными компонентами для поддержки коммуникации и обмена данными;
- **функциональные уровни**, включающие набор функциональных компонентов со схожими возможностями или служащими общей цели;
- **многоуровневые функции**, представляющие декомпозицию сложной функции на совокупность более простых; многоуровневые функции включают в себя функциональные компоненты с возможностями, используемыми на нескольких функциональных уровнях, и сгруппированы в подмножества.

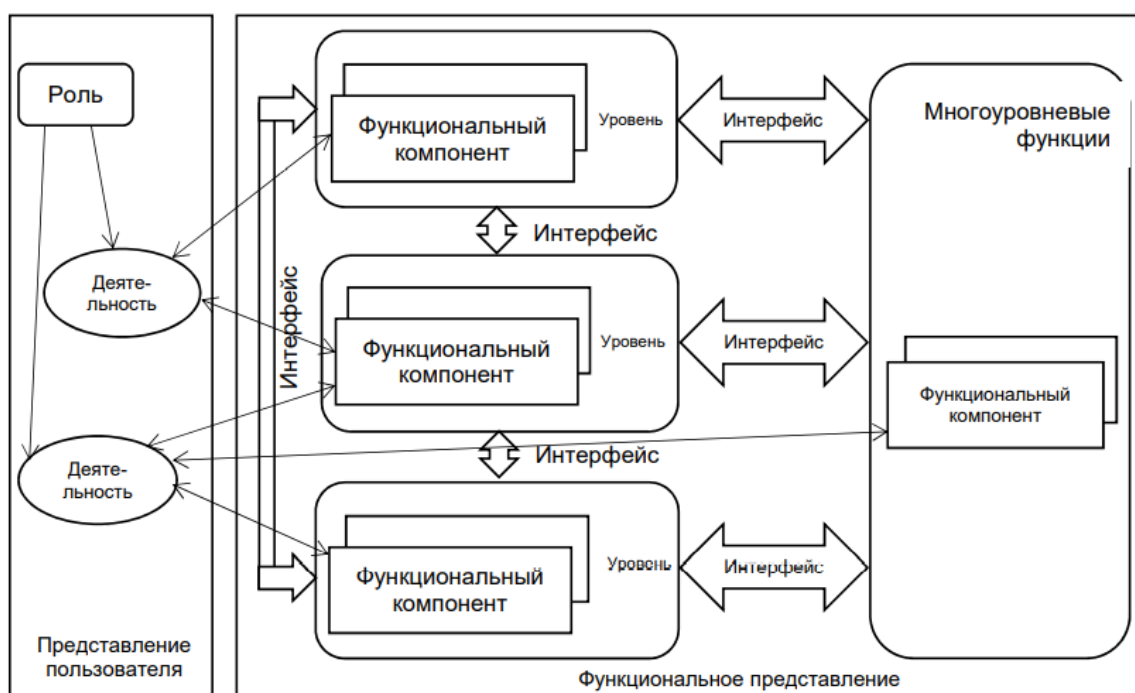


Рисунок 9 – Схема взаимодействия сущностей пользовательского и функционального представлений
Источник: [15]

Что касается сквозных аспектов, то они включают как архитектурные, так и эксплуатационные аспекты и применяются к компонентам и функциям эталонной архитектуры больших данных.

Сквозные аспекты влияют на деятельности с большими данными, выполняемые ролями. Роли, в свою очередь, могут координировать между собой функции поддержки сквозного аспекта. Каждому сквозному аспекту соответствуют функциональные компоненты для поддержки деятельностей и набор деятельностей с большими данными.

Сквозные аспекты являются общими для ролей, деятельностей и функциональных компонентов. Например, безопасность является сквозным аспектом, поскольку она применяется к сервис-провайдерам (доступа к большим данным, приложений больших данных, среды обработки больших данных), к потребителям больших данных и к партнёрам сервиса больших данных.

Применение эталонной архитектуры больших данных основано на реализации пошагового процесса, приведенного в разделе 4.

Таким образом, эталонная архитектура больших данных позволяет обеспечивать решение следующих задач стандартизации:

- выполнение формального описания различных компонентов, процессов и систем больших данных в контексте общей концептуальной модели больших данных;
- выполнение функциональной классификации, которая позволит заинтересованным государственным ведомствам, агентствам и другим потребителям классифицировать решения для больших данных и проводить их сравнительный анализ;
- проведение анализа стандартов по их функциональной совместимости.

7 Безопасность больших данных и конфиденциальность персональных данных

При применении эталонной архитектуры больших данных важное место занимает проблема обеспечения информационной безопасности, которая затрагивает все роли и подроли в её экосистеме и функциональных компонентах. Задачи обеспечения информационной безопасности отражены в четвертой части международного стандарта ISO/IEC 20547-4 [16], в котором представлено описание аспектов безопасности больших данных и их конфиденциальности в рамках эталонной архитектуры больших данных, включая роли, деятельности и функциональные компоненты.

Требуемый уровень информационной безопасности обеспечивается *оркестратором системы больших данных* при формировании политики информационной безопасности, обосновании требований к системе больших данных и проведении аудита информационной безопасности, а также *сервис-провайдерами* приложений больших данных и среды обработки больших данных при разработке системы, её развёртывании (вводе в эксплуатацию) и в процессе эксплуатации.

В стандарте ISO/IEC 20547-4 представлены результаты анализа проблем информационной безопасности, вытекающих из ключевых свойств больших данных и ключевых свойств процесса обработки данных, которые создают дополнительные риски и являются причинами снижения уровня безопасности и конфиденциальности при работе с большими данными, сформулированы рекомендации по обеспечению безопасности и конфиденциальности при выполнении различных операций с большими данными.

Парадигма больших данных привела к стиранию границы безопасности между системами сбора, хранения и доступа к данным — областями, которые традиционно рассматривались как независимые. В связи с этим первостепенное значение для создания атмосферы взаимного доверия и сотрудничества между заинтересованными сторонами в сфере сбора, хранения и обработки больших данных имеет стандартизация требований к безопасности и конфиденциальности персональных данных.

Ключевые свойства больших данных (объём, скорость обработки, разнообразие и изменчивость), а также ключевые свойства процесса обработки данных (волатильность, достоверность и ценность) определяют дополнительные риски и, следовательно, являются причинами снижения уровня безопасности и конфиденциальности при работе с большими данными:

- скорость обработки данных порождает риск, связанный с существенным повышением скорости потока, в рамках которого данные создаются, хранятся, анализируются или визуализируются; в этом случае средства управления безопасностью могут приводить к снижению скорости передачи данных, поэтому от них нередко отказываются;
- разнообразие данных значительно увеличивает их сложность, поскольку множество различных источников данных находится под контролем различных субъектов; повышение сложности данных неизбежно приводит к появлению новых уязвимостей; новые возможности, вызванные разнообразием больших данных, позволяют получить персональные данные из обезличенных наборов путем их сопоставления с доступными публичными базами данных;
- изменчивость данных влечет за собой риски, связанные с более быстрыми изменениями скорости передачи, формата/структуры, семантики и (или) качества данных; всё это приводит к необходимости совершенствования средств управления безопасностью в целях защиты персональных данных;

- волатильность данных негативно влияет на эффективность ведения журналов безопасности и усложняет управление безопасностью;
- необходимость обеспечения достоверности данных предъявляет более высокие требования к таким показателям, как целостность, согласованность и точность; возникающие при этом взаимосвязанные риски могут агрегироваться и значительно возрастать;
- ценность данных порождает большее число атак, преследующих различные цели и интересы.

Активное распространение приложений для работы большими данными вызывает всё более серьезные проблемы с точки зрения безопасности и конфиденциальности персональных данных, включая случаи их потери и утечки, а также скрытую неконтролируемую передачу данных. Это порождает злоупотребление данными, а также ставит под угрозу социальную стабильность и национальную безопасность.

В стандарте ISO/IEC 20547-4 значительное место отведено рассмотрению вопросов обеспечения безопасности и конфиденциальности персональных данных при использовании технологических платформ, в которых применяются разнообразные технологии обработки больших данных, новая техническая архитектура и вспомогательные платформы, а также специализированное программное обеспечение.

Выявлены недостатки существующих способов обеспечения информационной безопасности при применении в системах работы с большими данными и сформулированы следующие задачи совершенствования подходов к обеспечению их защиты.

А) Совершенствование традиционных средств управления безопасностью при работе с большими данными. Приложения для больших данных обычно используют открытую распределённую архитектуру вычислений и хранения со сложной базовой поддержкой для организации распределённых хранилищ больших данных и высокопроизводительных вычислительных сервисов. Благодаря этим новым технологиям и архитектурам границы приложений для больших данных размываются, поэтому традиционные методы защиты на основе границ перестают работать. Кроме того, целенаправленные устойчивые угрозы, распределённые атаки «отказ в обслуживании», интеллектуальный анализ данных на основе машинного обучения и способы обнаружения персональных данных, а также другие типы атак выявляют серьёзные недостатки традиционных инструментов защиты и обнаружения, а также других мер обеспечения безопасности. Необходимы новые технологические подходы для обеспечения конфиденциальности персональных данных, а также реализации методов машинного обучения, криптографических механизмов безопасности, ориентированных на данные, и средства контроля доступа.

Б) Обеспечение безопасности и конфиденциальности персональных данных при использовании инфраструктуры распределённых вычислений и хранения больших данных. С этой целью требуется разработка и внедрение инструментов безопасных распределённых вычислений и безопасных систем хранения данных. Для обработки больших данных необходимы масштабируемые и распределённые решения для безопасного хранения данных, проведения аудита и выявления источников происхождения данных. Аналитика в реальном времени для анализа угроз требует обработки больших объёмов данных, связанных с безопасностью.

В) Совершенствование механизмов обеспечения безопасности технологической платформы в контексте безопасности и конфиденциальности персональных данных при работе с большими данными. Многие существующие приложения для работы с большими данными используют платформы и технологии управления большими данными на базе платформы Hadoop и технологии обработки и программной модели для распределённых вычислений MapReduce.

На начальном этапе проектирования эти платформы и технологии в основном рассматриваются как решения для доверенной внутренней сети; вместе с тем функции аутентификации, авторизации, ключевые сервисы, а также возможности проведения аудита безопасности практически не учитываются. При этом общая эффективность средств обеспечения безопасности является недостаточной. Между тем в приложениях для больших данных часто используются сторонние компоненты с открытым исходным кодом. Из-за отсутствия надлежащего управления тестированием и сертификацией безопасности этих компонентов во многих случаях не

удается предотвратить появление уязвимостей и вредоносного программного обеспечения в приложениях для больших данных.

Г) Создание средств управления доступом к приложениям для работы с большими данными. Из-за большого разнообразия типов данных и широкого спектра приложений для работы с большими данными они часто используются для предоставления множества услуг пользователям с разными учётными данными и целями из разных организаций или разных подразделений одной организации. В общем случае контроль доступа является эффективным средством обеспечения контролируемого использования данных, однако из-за большого числа неизвестных пользователей и данных, к которым необходимо получить доступ, становится сложным предварительно определить роли и разрешения на доступ к данным. Права пользователя на доступ к данным могут быть классифицированы заранее, тем не менее многочисленность ролей снижает достоверность контроля фактических разрешений для каждой роли. Поэтому становится затруднительным для каждого пользователя точно определить диапазон данных, к которому ему необходимо получать доступ, без развёртывания более современной модели управления доступом.

Д) Создание масштабируемых механизмов обеспечения безопасности и конфиденциальности персональных данных. При разработке и применении механизмов обеспечения безопасности больших данных и конфиденциальности персональных данных, таких как управление ключами, управление идентификацией и доступом, обезличивание и т. д., в среде больших данных необходимо учитывать не только функции безопасности и конфиденциальности, но и масштабируемость этих механизмов, чтобы гарантировать обработку большого объёма данных, поступающих с высокой скоростью.

С точки зрения приложений для работы с данными с учётом их характеристик (объёма, разнообразия, вариативности, скорости обработки, достоверности, изменчивости), а также огромной ценности больших данных требуется решение следующих задач для обеспечения безопасности и конфиденциальности персональных данных.

- Создание средств защиты для обработки персональных данных в рамках системы больших данных, которые должны учитывать особенности распределённых систем, открытой сетевой среды, сложных приложений для обработки данных.
- Создание средств обеспечения конфиденциальности персональных данных при работе с большими данными с учётом таких инцидентов безопасности, как злоупотребление данными, внутренняя кража и сетевые атаки, ведущие к утечке персональных данных, которые будут иметь более серьёзные последствия, чем в обычных информационных системах.
- Создание средств контроля аутентичности данных и источников больших данных с учётом их широкого спектра, включая различные датчики, активные загрузки и общедоступные веб-сайты, а также большое количество ненадёжных источников, через которые злоумышленники имеют возможность фальсифицировать данные. При проверке аутентичности данных возникает множество проблем из-за ограниченной производительности терминалов сбора данных, отсутствия необходимых технологий, лимитированного объёма информации, а также разнообразия и сложности источников.
- Создание средств защиты прав владельцев данных при работе с большими данными с учётом того, что во время работы с большими данными к ним могут обращаться различные пользователи, их могут передавать от одного специалиста к другому, а также анализировать для получения новых данных. То есть в процессе обмена данными и общего доступа к ним возникает обстоятельство, при котором право владения данными для владельца данных и право использования данных разделяются. Иными словами, данные могут находиться вне контроля владельца, что влечёт за собой риски злоупотребления данными, неопределённого владения ими и неясных требований по надзору за их безопасностью, что может нанести серьёзный ущерб правам и интересам владельцев данных.

Поскольку экосистемы больших данных и сети организаций обмениваются данными с целью сбора, обработки, хранения и анализа, должны быть предусмотрены следующие варианты взаимодействия между заинтересованными сторонами:

- для согласования общих требований к безопасности и защите персональных данных в экосистеме и соответствующих требований в отдельной организации;

- для согласования вопросов общего управления рисками на уровне экосистемы и уровне отдельной организации;
- для обеспечения того, чтобы отдельные организации гарантировали согласованную работу с защищаемыми активами.

В стандарте представлены аспекты безопасности и защиты персональных данных с точки зрения пользователя и с точки зрения мероприятий по верхнеуровневому управлению организацией в пространствах проблем и решений. При этом пространство проблем относится к миру проблем и мотиваций конечного пользователя, а пространство решений – к миру продуктов, услуг и технологий.

В стандарте ISO/IEC 20547-4 также представлены рекомендации по обеспечению безопасности и защите персональных данных при работе с большими данными как на *уровне организации, так и на уровне экосистемы больших данных.*

Взаимосвязи между организациями, экосистемой больших данных, а также их эталонной архитектурой при решении задач обеспечения безопасности и конфиденциальности персональных данных показаны на рисунке 10.

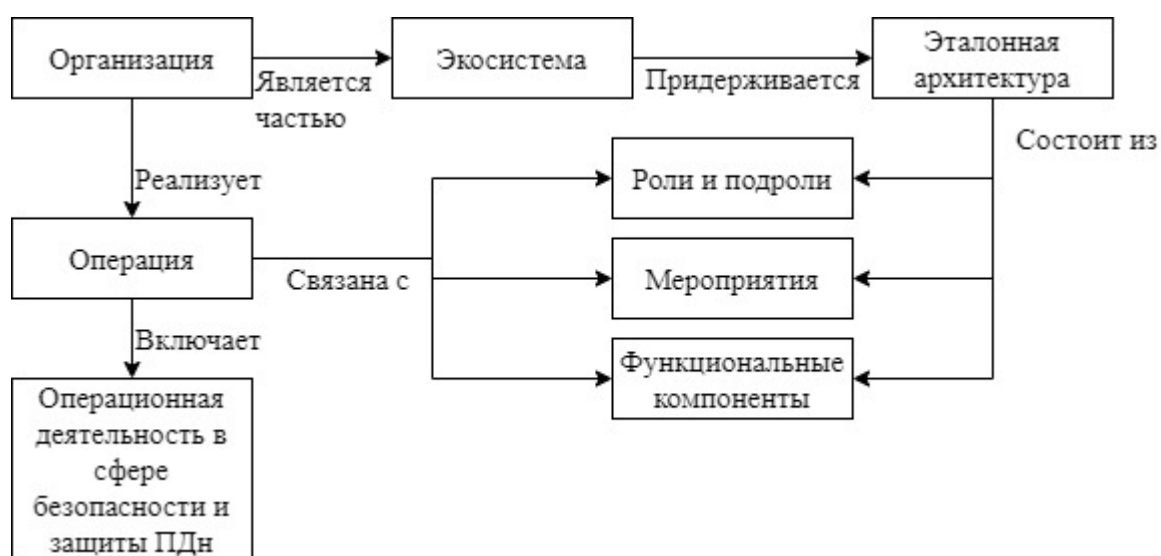


Рисунок 10 – Структурная схема взаимосвязи между организациями, экосистемой больших данных и эталонной архитектурой больших данных
Источник: [16]

Организация может быть частью экосистемы, которая придерживается принципов эталонной архитектуры больших данных. Эталонная архитектура больших данных включает роли и подроли, деятельности и функциональные компоненты.

Организация осуществляет операции, связанные с ролями, деятельностями и функциональными компонентами. Например, организация играет роль сервис-провайдера приложений для больших данных. Организация осуществляет операции, ориентированные на безопасность и конфиденциальность персональных данных. Они применяются для защиты активов от уязвимостей. Некоторые активы могут быть специфическими для организации, например сведения, содержащие коммерческую тайну, в то время как другие могут совместно использоваться в экосистемах, например наборы данных.

Подход, представленный в стандарте ISO/IEC 20547-4, предусматривает описание проблем информационной безопасности с точки зрения двух типов операций в области обеспечения безопасности и конфиденциальности персональных данных: операции по организации деятельности по обеспечению информационной безопасности и защите персональных данных, а также операции по сотрудничеству в экосистеме.

В стандарте представлены рекомендации по обеспечению безопасности и защите персональных данных при работе с большими данными на уровне организации и на уровне экосистемы, которые соответствуют этапам: разработки требований; анализа рисков;

проектирования средств управления; разработки функционала системы для работы с большими данными.

Применительно к каждому этапу сформулированы особенности решения задач обеспечения безопасности и защиты персональных данных при работе с большими данными. В основу рекомендаций положено описание операционной деятельности, представленное в ряде стандартов, посвященных управлению информационной безопасностью, защите персональных данных и проектированию систем защиты персональных данных [ИСО/МЭК 27001 Системы управления информационной безопасностью. Требования; ISO/IEC 27000 Применение стандарта ISO/IEC 27001 в конкретных секторах. требования; ИСО/МЭК 29100 Основы защиты персональных данных; ИСО/МЭК 27701 Дополнение к ISO/IEC 27001 и ISO/IEC 27002 для управления информацией о защите персональных данных; ИСО/МЭК 27005 Управление рисками, связанными с информационной безопасностью; ИСО/МЭК 27550 Проектирование систем защиты персональных данных и др.]

Стандарт ISO/IEC 20547–4 содержит описание *функциональных компонентов* безопасности и конфиденциальности персональных данных, под которыми понимаются функциональные элементы эталонной архитектуры, используемые для выполнения определенной процедуры или совокупности процедур безопасности в конкретной реализации архитектуры. В качестве функциональных компонентов могут использоваться категории управления безопасностью, представленные в национальном стандарте ГОСТ Р ИСО/МЭК 27002 [60].

Таким образом, содержание стандарта ISO/IEC 20547–4 включает описание различных аспектов обеспечения безопасности больших данных и конфиденциальности персональных данных применительно к эталонной архитектуре, в том числе задач совершенствования существующих подходов, а также рекомендации по обеспечению безопасности и конфиденциальности при выполнении операций с большими данными в отдельной организации, а также в экосистеме больших данных.

8 Стандартизация требований заказчика к действиям, связанным с оперированием большими данными

Система международных стандартов ISO/IEC 20546 / 20547-X формирует системно-технологическое окружение оперирования большими данными – эталонную архитектуру, виды данных и технологии их обработки, хранения, анализа, многое другое.

Однако парадигма «больших данных» является значительно более ёмкой и не ограничивается большими массивами данных и системами для работы с ними. Наряду с проблемами описания системно-технологического окружения при использовании больших данных не менее существенными являются вопросы упорядочения взаимоотношений между заказчиками и их исполнителями (подрядчиками, поставщиками) при оперировании большими данными. Для заказчика основным становится удовлетворение его потребностей, которые должны быть сформулированы в виде однозначных, полных требований к результатам действий с большими данными и быть понятны исполнителю. Как правило, такие требования фиксируются в документе, представляющем собой техническое задание, регламентирующее взаимоотношения заказчика и исполнителя, а также видение заказчиком ожидаемого результата.

Разработанный в 2020–2021 годах проект национального стандарта «Информационные технологии. Большие данные. Техническое задание. Требования к содержанию и оформлению» [40] направлен на типизацию и стандартизацию требований, включаемых в технические задания для следующих видов деятельности, связанной с большими данными:

- поставка массивов (наборов) больших данных;
- оперирование большими данными;
- разработка и/или поставка технологий и/или средств (решений) для оперирования большими данными;
- внедрение технологий и/или средств (решений) для оперирования большими данными;
- предоставление в пользование (аренда) технологий и/или средств (решений) оперирования большими данными;
- техническая эксплуатация средств (решений) для оперирования большими данными.

При разработке национального стандарта в значительной мере использованы термины и определения, установленные международными стандартами и другими национальными стандартами. В то же время возникла необходимость в уточнении или введении новых понятий для типизации субъектов процессов оперирования большими данными или для описания взаимоотношений между ними. Например, под терминами «заказчик», «исполнитель (подрядчик)» и «поставщик» в стандарте понимаются лица, которые могут заказать, дать задание или поручение, тем самым определяя свои потребности, и, соответственно, лица, которые могут эти заказы, задания или поручения выполнять для удовлетворения потребностей заказчика. Эти лица могут выступать как в роли «внешних», так и «внутренних» заказчиков, исполнителей (подрядчиков) или поставщиков.

Общие положения стандарта предусматривают:

- определение техническим заданием предмета выполняемых работ (оказываемых услуг) и/или поставок, а также основные требования к ним и их результатам;
- использование в качестве основы для технического задания существующей и прогнозируемой потребности заказчика, результатов исследований проблем, связанных с оперированием большими данными, национальных, международных и отраслевых стандартов и иных нормативных технических документов, требований, установленных нормативными правовыми актами, опыта предыдущих аналогичных выполненных работ (оказанных услуг), выполненных поставок;
- исключение возможности различных толкований как содержания технического задания в целом, так и устанавливаемых им требований в частности;
- обеспечение логической связанности содержания технического задания, достаточной для понимания целей, задач и требований к выполнению работ (оказанию услуг), осуществлению поставок и ожидаемым результатам;
- наличие положений о проверке соответствия выполнения работ (оказания услуг), осуществления поставок и их результатов установленным техническим заданием целям, задачам и требованиям. Еще одной особенностью национального стандарта является разделение видов деятельности, направленных на непосредственное использование больших данных по назначению и на обеспечение такого использования. Ко второй категории относятся поставка массивов (наборов) больших данных, разработка и/или поставка технологий и/или средств (решений) для оперирования большими данными, внедрение технологий и/или средств (решений) для оперирования большими данными, предоставление в пользование (аренда) технологий и/или средств (решений) оперирования большими данными, техническая эксплуатация средств (решений) для оперирования большими данными. Подобный подход позволяет устанавливать и детализировать специфические требования, относящиеся к каждому из перечисленных видов деятельности.

Разработанный стандарт не ограничивает возможности включения в задание произвольной комбинации различных видов деятельности, в том числе оперирования данными и деятельности, направленной на обеспечение такого использования. Предусмотрена возможность формирования технического задания для оперирования большими данными в рамках всего «жизненного цикла данных».

В отдельном разделе стандарта определен порядок формирования технических заданий по осуществлению деятельности, связанной с использованием больших данных для государственных и муниципальных нужд.

9 Качество данных

Обеспечение единообразных требований к качеству данных имеет первостепенное значение для эффективной работы с большими данными и достижения полезных результатов в современной динамике глобальной цифровой экономики. Первыми результатами работ по стандартизации требований к качеству данных стали международные и национальные стандарты качества данных [61, 62] и серия стандартов ISO 9000 [63], а также разрабатываемые в настоящее время проекты международных стандартов ISO/IEC 5259-X «Искусственный интеллект. Качество данных для аналитики и машинного обучения» [18–21].

Понятие качества данных вобрало в себя смыслы двух понятий: качество и данные. Апогей стандартизованного формирования понятия «качество» можно отнести к построению системы менеджмента качества в условиях рыночной экономики. Феномен качества наиболее рельефно выражен в серии стандартов ISO 9000. Качество – это определённое явление рыночной экономики, сочетающее в себе два процесса: рост удовлетворённости потребителей продукции при постоянном снижении её себестоимости (издержек). Содержательный интерес может представлять понимание термина «качество», как синтеза двух явлений: аналитики и романтики [64]. Первое явление предполагает возможность своей формализованной репрезентации, в том числе, с применением различных методов математики, физики, инструментов визуализации и искусственного интеллекта. Второе явление отрицает возможность его непосредственной формализации, оно больше ассоциируется с чисто субъективными особенностями человека, такими как: эмоции, мысли, чувства, интуиция, свобода воли и др.

С развитием тренда больших данных обострился вопрос качества данных. Это связано с необходимостью высококачественного поиска информации в больших массивах данных, и, что более важно, использования данных для обучения нейронных сетей, тестирования моделей искусственного интеллекта. Обеспечение качества данных становится всё более сложным процессом, поскольку в данных содержится много нерелевантных данных, шума, мусора. Неслучайно появляются результаты исследований, название которых красноречиво подтверждают актуальность темы качества данных, а именно, «тёмные данные» [65], «грязные данные» [66]. Например, к грязным данным относят данные, содержащие ошибочную информацию. При этом полное удаление грязных данных из источника нецелесообразно или практически невозможно. В этой связи всё более актуализируется вопрос очистки данных, см., например, работу [67]. В этих работах акцентируется внимание на таких аспектах, как:

- известные данные в массиве отсутствуют, в данных есть пробелы;
- пользователь может даже не знать, что ему не хватает данных;
- неправильный выбор критериев для включения в выборку;
- самовыбор, когда люди сами решают, стоит ли данные включать в базу данных;
- неправильная оценка критических аспектов системы для обработки данных;
- возможность предположения иных данных, если бы были предприняты другие действия;
- данные могут быть неверными, неточными, дублировать друг друга и вводить в заблуждение;
- данные могут быть неинтегрируемыми и не допускать обобщённого форматирования;
- изменение данных и их значений во времени и др.

С ростом объёма данных всё более актуальным становится вопрос их семантической интерпретации. Если раньше истинность, и, соответственно, качество данных обычно оценивалась через их отображение на объекты или артефакты реальной действительности, то развитие технологий искусственного интеллекта потребовало всё больше внимания уделять неформализуемым семантическим интерпретациям моделей, так называемым, когнитивным семантикам [68]. Они отражают мыслительные, эмоциональные и трансцендентальные аспекты индивидуального и коллективного сознания, феномены сознательного и бессознательного. Повидимому, охват именно таких семантик будет лежать в основе перспективного развития методов искусственного интеллекта.

Интерес в контексте качества данных может представлять процесс потери сведений и информации при преобразовании и фильтрации исходных аналоговых данных в цифровую форму [69]. Например, это может происходить при обработке сигналов, получаемых из космоса или с ускорителей квантовых частиц.

Таким образом, современные стандарты, направленные на обеспечение высокого качества данных, можно оценить на соответствие следующим критериям:

- соответствие стандартам системы менеджмента качества;
- полнота охвата феномена «данные» в контексте понятий «информация» и «сведения»;
- учёт формализуемых (денотативных) и неформализуемых (когнитивных) семантик;
- взаимосвязь данных с инструментами их обработки и анализа, например, глубокого обучения;
- минимизация искажений при трансформации исходных аналоговых данных в цифровую форму;
- учёт нарастающего объёма «тёмных» данных;

- учёт построения систем очистки и сохранения данных.

Указанные международные и национальные стандарты формируют систему понятий относительно феномена качества данных, определяют характеристики качества данных, устанавливают к ним требования и указывают способы повышения качества информации.

Понятие «качество» в рассматриваемых стандартах задаётся согласно стандартам системы менеджмента качества, а именно, это степень соответствия совокупности присущих характеристик объекта требованиям. Однако в сочетании с термином «данные» соответствие новой коллокации остальным критериям вызывает сомнение.

Так, определение термина «данные (data)» в серии стандартов ISO 9000 механически заимствовано из иного стандарта (ISO/IEC 2382:2015 [70]). Оно формулируется как «интерпретируемое представление информации в соответствующей форме, удобной для передачи, интерпретации и обработки». Такое заимствование этого термина из другого стандарта в контексте темы качества данных далеко не в полной мере отражает аспект возможной семантической интерпретации данных несмотря на то, что термин «интерпретация» в этом коротком определении встречается дважды. То есть в определении термина «качество данных» и самом стандарте умаляется учёт формализуемых и неформализуемых особенностей семантической интерпретации данных, которая может быть как денотативной, так и когнитивной. Рассматриваемые стандарты далеко не в полной мере отвечают и другим перечисленным выше критериям соответствия.

Определение термина «качество данных» в проектах серии стандартов ISO 5259-X осуществляется через задание соответствующих характеристик на основе международного стандарта ISO/IEC 25012:2008 [62] и демонстрацией взаимосвязи между моделью качества данных, требованиями к качеству данных для аналитики и машинного обучения, а также характеристиками качества данных. Рассматривается полнота и применимость для аналитики и машинного обучения следующих пятнадцати характеристик качества данных: аккуратность, полнота, согласованность, надёжность, правильность, доступность, соответствие, конфиденциальность, эффективность, точность, прослеживаемость, понятность, доступность, переносимость и возможность восстановления.

Таким образом, анализ разрабатываемых и разработанных базовых стандартов искусственного интеллекта в части адекватности определения термина «качества данных» требованиям реальной теории и практики позволяет сделать следующие выводы относительно этого определения:

- проводится в основном механическое сложение определений двух терминов, приведённых в других стандартах, что не позволяет обеспечить должную синергию термина и его понятия;
- не учитываются различные виды семантической интерпретации данных, что затрудняет перспективное развитие систем обработки данных, в том числе с применением искусственного интеллекта;
- в уточнении и определении нуждаются учёт в стандарте нарастающего объёма ошибок данных и процедуры трансформации исходных аналоговых данных в цифровую форму.

Заключение

Обзор процесса и содержания стандартизации работы с большими данными демонстрирует его актуальность и значимость для социально-экономического развития. В Российской Федерации за последние годы резко сократилось отставание от международных процессов стандартизации работы с большими данными. Стандарты работы с большими данными в первую очередь будут востребованы органами власти и коммерческими компаниями, которые решают управленческие задачи или ведут бизнес, принимая решения на основе данных. Следует отметить, что введение в действие серии национальных стандартов, аналогичных международным стандартам серии 20547-X, имеет важное значение в реализации задач Национальной программы «Цифровая экономика Российской Федерации», в том числе в части внедрения сквозной цифровой технологии «Большие данные».

Очевидно, что одним из наиболее перспективных и направлений дальнейшей стандартизации больших данных станет разработка стандартов, фиксирующих требования к качеству массивов данных для разнообразных вариантов использования и различных технологий,

начиная от традиционной аналитики больших данных и заканчивая технологиями машинного/глубокого обучения. Данная тенденция является логическим развитием процесса стандартизации, первый этап которого завершился утверждением международных общепромышленных стандартов эталонной архитектуры больших данных.

Вторым важным направлением станет разработка «отраслевых» стандартов эталонных архитектур систем для работы с большими данными применительно к конкретным сферам деятельности, таким как телекоммуникации, здравоохранение, образование или транспорт.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. ISO/IEC 20546:2019 Information technology – Big data – Overview and vocabulary // International Organization for Standardization. URL: <https://www.iso.org/standard/68305.html> (дата обращения: 01.10.2021).
2. ГОСТ Р ИСО/МЭК 20546–2021 «Информационные технологии. Большие данные. Обзор и словарь» // Российский институт стандартизации. URL: <https://www.gostinfo.ru/catalog/Details/?id=6859575> (дата обращения: 01.10.2021).
3. ISO/IEC DIS 22989 Information technology – Artificial intelligence – Artificial intelligence concepts and terminology // International Organization for Standardization. URL: <https://www.iso.org/standard/74296.html> (дата обращения: 01.10.2021).
4. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh Ch., Byers A.H. Big data: The next frontier for innovation, competition, and productivity // McKinsey Global Institute. P. 12. URL: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#> (дата обращения: 01.10.2021).
5. Big Data interoperability Framework. V1.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V1_output_docs.php (дата обращения: 01.10.2021).
6. Big Data interoperability Framework. V2.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V2_output_docs.php (дата обращения: 01.10.2021).
7. Big Data interoperability Framework. V3.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V3_output_docs.php (дата обращения: 01.10.2021).
8. Big Data: Big today, normal tomorrow. ITU T Technology Watch Report. November 2013. // International Telecommunication Union. URL: https://www.itu.int/en/ITU-T/techwatch/Pages/big_data_standards.aspx (дата обращения: 01.10.2021).
9. Big data: Preliminary Report 2014. ISO/IEC JTC1, 2015. // International Organization for Standardization. URL: https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/big_data_report-jtc1.pdf (дата обращения: 01.10.2021).
10. ITU-T Y.3600 (11/2015) Big data – Cloud computing-based requirements and capabilities. // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=12584&lang=en> (дата обращения: 01.10.2021).
11. ITU-T Y Suppl. 40 (07/2016) Big data standardization roadmap // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=13022&lang=en> (дата обращения: 01.10.2021).

12. ISO/IEC TR 20547-2:2018 Information technology – Big data reference architecture – Part 2: Use cases and derived requirements // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-2:ed-1:v1:en> (дата обращения: 01.10.2021).
13. ISO/IEC TR 20547-5:2018 Information technology – Big data reference architecture – Part 5: Standards roadmap // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-5:ed-1:v1:en> (дата обращения: 01.10.2021).
14. ISO/IEC TR 20547-1:2020 Information technology – Big data reference architecture – Part 1: Framework and application process // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-1:ed-1:v1:en> (дата обращения: 01.10.2021).
15. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20547:-3:ed-1:v1:en> (дата обращения: 01.10.2021).
16. ISO/IEC 20547-4:2020 Information technology – Big data reference architecture – Part 4: Security and privacy // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20547:-4:ed-1:v1:en> (дата обращения: 01.10.2021).
17. Walshe R. The Road to Big Data Standardisation // The Elements of Big Data Value / E. Curry et al. (eds.). Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-68176-0_14.
18. ISO/IEC WD 5259-1 Data quality for analytics and ML – Part 1: Overview, terminology, and examples // International Organization for Standardization. URL: <https://www.iso.org/standard/81088.html> (дата обращения: 01.10.2021).
19. ISO/IEC AWI 5259-2 Data quality for analytics and ML – Part 2: Data quality measures // International Organization for Standardization. URL: <https://www.iso.org/standard/81860.html> (дата обращения: 01.10.2021).
20. ISO/IEC WD 5259-3 Data quality for analytics and ML – Part 3: Data quality management requirements and guidelines // International Organization for Standardization. URL: <https://www.iso.org/standard/81092.html> (дата обращения: 01.10.2021).
21. ISO/IEC WD 5259-4 Data quality for analytics and ML – Part 4: Data quality process framework // International Organization for Standardization. URL: <https://www.iso.org/standard/81093.html> (дата обращения: 01.10.2021).
22. ITU-T Y.3603 Big data – Requirements and conceptual model of metadata for data catalogue» // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14137&lang=en> (дата обращения: 01.10.2021).
23. ITU-T Y.3604 Big data – Overview and requirements for data preservation» // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14138&lang=en> (дата обращения: 01.10.2021).
24. Resolution 166. Registration of a PWI entitled «Information technology – Artificial intelligence – Data life cycle framework». ISO/IEC JTC 1/SC 42 “Artificial intelligence”. Resolutions taken during the Closing Plenary - JTC 1/SC 42 - 7 May 2021.
25. ITU-T Y.3651 Big-data-driven networking – mobile network traffic management and planning // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=13818&lang=en> (дата обращения: 01.10.2021).
26. ITU-T Y.3653 Big data driven networking – functional architecture // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14615&lang=en> (дата обращения: 01.10.2021).
27. Приказ Федерального агентства по техническому регулированию и метрологии (Росстандарт) от 25.07.2019 № 1732 (ред. от 20.01.2021) «О создании технического комитета по стандартизации "Искусственный интеллект» // Росстандарт. URL: <https://www.rst.gov.ru/portal/gost/home/activity/documents/orders#/order/104460> (дата обращения: 01.10.2021).
28. Приказ Федерального агентства по техническому регулированию и метрологии (Росстандарт) от 20.08.2020 № 1415 «О внесении изменений в приказ Федерального агентства по техническому регулированию и метрологии от 25 июля 2019 г. № 1732 «О создании технического комитета по стандартизации "Искусственный интеллект» // Росстандарт. URL:

- <https://www.gost.ru/portal/gost/home/activity/documents/orders#/order/179415> (дата обращения: 01.10.2021).
29. Распоряжение Правительства РФ от 23.03.2018 N 482-р (ред. от 28.05.2020) «Об утверждении плана мероприятий («дорожной карты») по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации Национальной технологической инициативы по направлению «Технет» (передовые производственные технологии)» // СЗ РФ. 2018. № 15 (ч. V). Ст. 2173.
 30. Российская венчурная компания. URL: https://www.rvc.ru/upload/iblock/ac9/Long-term_plan_of_standardization_Technet.pdf (дата обращения: 01.10.2021).
 31. Приказ Росстандарта от 05.02.2019 г. № 166 «О внесении изменений в Программу национальной стандартизации на 2019 год, утверждённую приказом Федерального агентства по техническому регулированию и метрологии от 1 ноября 2018 г. № 2285» // Росстандарт. URL: <https://www.gost.ru/portal/gost/home/activity/standardization> (дата обращения: 01.10.2021).
 32. Протокол результатов общественного обсуждения эффективности применения принятых во исполнение планов мероприятий «дорожных карт» по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации Национальной технологической инициативы нормативных правовых актов и документов по стандартизации, достигнутых целей их применения не менее чем в течение одного года с даты начала применения (реализации) акта // Национальная технологическая инициатива. URL: <https://nti2035.ru/upload/351708.2.protocol.pdf> дата обращения: 01.10.2021).
 33. Дорожная карта развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект» // Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. URL: <https://digital.gov.ru/ru/documents/6658/> (дата обращения: 01.10.2021).
 34. Паспорт Федерального проекта «Нормативное регулирование цифровой среды» // Министерство экономического развития Российской Федерации. URL: https://www.economy.gov.ru/material/directions/gosudarstvennoe_upravlenie/normativnoe_regulirovanie_cifrovoy_sredy/ (дата обращения: 01.10.2021).
 35. Национальная стратегия развития искусственного интеллекта на период до 2030 года, утверждённая Указом Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» // СЗ РФ. 2019. № 41. Ст. 5700.
 36. Федеральный проект «Искусственный интеллект» // Министерство экономического развития Российской Федерации. URL: https://www.economy.gov.ru/material/directions/tehnologicheskoe_razvitie/federalnyy_proekt_iskusstvennyy_intellekt/ (дата обращения: 01.10.2021).
 37. Перспективная программа стандартизации по приоритетному направлению «искусственный интеллект» на период 2021-2024 годы, утверждённая Министерством экономического развития Российской Федерации и Федеральным агентством по техническому регулированию и метрологии 22.12.2020 г. // URL: https://www.economy.gov.ru/material/news/v_rossii_poyavyatsya_standarty_v_oblasti_iskusstvennogo_intellekta.html (дата обращения: 01.10.2021).
 38. Окончательная редакция проекта ГОСТ Р «Информационные технологии. Эталонная архитектура больших данных. Часть 1. Структура и процесс применения» // Технический комитет 164 «Искусственный интеллект». URL: <https://www.tc164.ru/окончательные-редакции> (дата обращения: 01.10.2021).
 39. Окончательная редакция проекта ГОСТ-Р «Информационные технологии. Эталонная архитектура больших данных. Часть 2. Варианты использования и производные требования» // Технический комитет 164 «Искусственный интеллект». URL: <https://www.tc164.ru/окончательные-редакции> (дата обращения: 01.10.2021).
 40. Проект ПНСТ «Информационные технологии. Большие данные. Типовая архитектура» // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/554715621> (дата обращения: 01.10.2021).
 41. Первая редакция проекта ГОСТ Р «Информационные технологии. Эталонная архитектура больших данных. Часть 5. Направления стандартизации» // Технические комитет 164

- «Искусственный интеллект». URL: <https://www.tc164.ru/первые-редакции> (дата обращения: 01.10.2021).
42. Окончательная редакция проекта ГОСТ Р «Информационные технологии. Большие данные. Техническое задание. Требования к содержанию и оформлению» // Технический комитет 164 «Искусственный интеллект». URL: <https://www.tc164.ru/окончательные-редакции> (дата обращения: 01.10.2021).
 43. Первая редакция проекта ГОСТ Р «Информационные технологии – Искусственный интеллект – Структура управления процессами аналитики больших данных» // Технический комитет 164 «Искусственный интеллект». URL: <https://www.tc164.ru/первые-редакции> (дата обращения: 01.10.2021).
 44. ISO/IEC 24668 Information technology – Artificial intelligence – Process management framework for big data analytics // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:24668:dis:ed-1:v1:en> (дата обращения: 01.10.2021).
 45. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis // Scientometrics. 2020. Vol 122. P. 1563–1581. <https://doi.org/10.1007/s11192-020-03371-2>.
 46. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
 47. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends // J Big Data. 2017. Vol 4, № 30. <https://doi.org/10.1186/s40537-017-0088-1>.
 48. de Meer, J. A Theory on Big Data. // INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge) / Draude, C., Lange, M. & Sick, B. (Hrsg.). Bonn: Gesellschaft für Informatik e.V. 2019. S. 261–269. https://doi.org/10.18420/inf2019_ws30.
 49. Walshe R., Casey K., Kernan J., Fitzpatrick D. AI and Big Data Standardization: Contributing to United Nations Sustainable Development Goals // Journal of ICT Standardization: 2020: Vol 8 Iss 2. URL: <https://journals.riverpublishers.com/index.php/IJCTS/article/view/2649> (дата обращения: 01.10.2021).
 50. Caballero I., Parody L., Bermejo I., Gómez López M.T., Gasca R.M., Piattini M. Service level agreement for data quality governed by Iso 8000-1X0 // Proceedings of the 19th International Conference on Information Quality, ICIQ. 2014. P. 114–127.
 51. Xiao Y., Lu L.Y.Y., Liu J.S., Zhou Z. Knowledge diffusion path analysis of data quality literature: A main path analysis // Journal of Informetrics. 2014. Vol 8 Iss 3. P. 594–605. <https://doi.org/10.1016/j.joi.2014.05.001>.
 52. Perin U. Reference Architectures and Standards for the Internet of Things and Big Data in Smart Manufacturing // International Journal of Recent Technology and Engineering. 2019. Vol 8 Iss 1C2. P. 884–887. <https://doi.org/10.1109/FiCloud.2019.00041>.
 53. Heale B.S.E., Overby C.L., Del Fiol G. et al. Integrating genomic resources with electronic health records using the HL7 infobutton standard // Applied Clinical Informatics. 2016. Vol 7 Iss 3. P. 817–831. <https://doi.org/10.4338/ACI-2016-04-RA-0058>.
 54. Raghunath N. A Standard for Benchmarking Big Data Systems // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. Vol 8585. P. 193–201. https://doi.org/10.1007/978-3-319-10596-3_15.
 55. Laney, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Research Note, 6, February 2001. URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (дата обращения: 01.10.2021).
 56. Cloutier R., Muller G., Verma D., Nilchiani R., Hole E., Bone M. The Concept of Reference Architecture // Systems Engineering. 2009. Vol 13, Iss 1. P 14–27. <https://doi.org/10.1002/sys.20129>.
 57. Big data: The next frontier for innovation, competition, and productivity. Report // McKinsey Global Institute. 2011. May. URL: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (дата обращения: 01.10.2021).
 58. Mohd Rehan Ghazi Durgaprasad Gangodkar Hadoop, MapReduce and HDFS: A Developers Perspective // Procedia Computer Science. 2015. Vol 48. P. 45–50.

- <https://doi.org/10.1016/j.procs.2015.04.108>. URL:
<https://www.sciencedirect.com/science/article/pii/S1877050915006171> (дата обращения: 01.10.2021).
59. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters // OSDI'04: Sixth Symposium on Operating System Design and Implementation. San Francisco, CA, 2004. P. 137-150. URL: <https://static.googleusercontent.com/media/research.google.com/ru//archive/mapreduce-osdi04.pdf> (дата обращения: 01.10.2021).
 60. ГОСТ Р ИСО/МЭК 27002–2021 Информационные технологии. Методы и средства обеспечения безопасности. Свод норм и правил применения мер обеспечения информационной безопасности // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/1200179669> (дата обращения: 01.10.2021).
 61. ГОСТ Р 56214–2014/ISO/TS 8000–1:2011 Качество данных. Часть 1. Обзор // Электронный фонд правовых и нормативно-технических документов. URL: <https://docs.cntd.ru/document/1200114769?section=text> (дата обращения: 01.10.2021).
 62. ISO/IEC 25012:2008 Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:25012:ed-1:v1:en> (дата обращения: 01.10.2021).
 63. ISO 9000:2015 Quality management systems – Fundamentals and vocabulary // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:ru> (дата обращения: 01.10.2021).
 64. Пирсиг Р. Дзэн и искусство ухода за мотоциклом/ пер. с англ. М.Горшкова. – СПб.: «Симпозиум», 2002.
 65. David J. Hand. Dark Data. Princeton University Press. Princeton and Oxford., 2020.
 66. Dirty Data // Techopedia. URL: <https://www.techopedia.com/definition/1194/dirty-data> (дата обращения: 01.10.2021).
 67. Raikov A.N., Avdeeva Z., and Ermakov A. Big Data Refining on the Base of Cognitive Modeling // Proceedings of the 1st IFAC Conference on Cyber-Physical&Human-Systems, Florianopolis, Brazil. 2016. P. 147-152. <https://doi.org/10.1016/j.ifacol.2016.12.205>.
 68. Raikov A. Cognitive Semantics of Artificial Intelligence: A New Perspective // Springer Singapore, Topics: Computational Intelligence XVII, 2021. <https://doi.org/10.1007/978-981-33-6750-0>.
 69. Райков А.Н., Котельников В.А. Предвосхищение будущего цифровизации // Информатизация и связь. 2019. № 1. – С. 7–11. <https://doi.org/10.34219/2078-8320-2019-10-1-7-11>.
 70. ISO/IEC 2382:2015 Information technology – Vocabulary // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en> (дата обращения: 01.10.2021).

BIG DATA STANDARDIZATION: INTERNATIONAL AND NATIONAL STANDARDS

Averkin, Alexei Nikolaevich

*Candidate of physical and mathematical sciences, associate professor
Plekhanov Russian University of Economics, Educational and scientific laboratory for artificial intelligence,
neurotechnology and business analytics, leading researcher
Moscow, Russia
averkin2003@inbox.ru*

Afanasev, Sergei Dmitrievich

*Candidate of law sciences
Lomonosov Moscow State University, National center for digital economy, lead specialist
Moscow Region State University, Institute of economics, management and law, Faculty of law, Department of
constitutional and municipal law, associate professor
Moscow, Russian Federation
sergei.afanasev@digital.msu.ru*

Mikryukov, Andrey Alexandrovich

*Candidate of engineering sciences, associate professor
Plekhanov Russian University of Economics, Institute of mathematics, information systems and digital economy,
Department of applied informatics and information security, associate professor
Moscow, Russian Federation
mikryukov.aa@rea.ru*

Padzhev, Valentin Valentinovich

*Institute of the Information Society, head of Directorate of law
Moscow, Russian Federation
vpadzhev@iis.ru*

Raikov, Alexander Nikolaevich

*Doctor of engineering sciences, professor
Lomonosov Moscow State University, National center for digital economy, head of Department for intellectual
technologies
V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, senior researcher
Moscow, Russian Federation
Alexander.N.Raikov@gmail.com*

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Khramtsovskaya, Natalya Alexandrovna

*Candidate of historical sciences
Electronic Office Systems LLC, leading expert in documentation management
Moscow, Russian Federation
sspchram@tochka.ru*

Abstract

The standards that establish requirements for big data technologies are designed to increase the efficiency of using these technologies in various sectors of the economy. The article examines the features of international and Russian national industry-wide data standards based on ISO/IEC 20546, 20547-X, 9000 series and 5259-X series projects in terms of basic concepts, reference architecture and requirements for big data and their quality, as well as approaches

to standardizing customer requirements for actions related to the use of big data. It is concluded that the development of documents on big data standardization in Russia is of great importance and entails the need to increase the pace of standardization.

Keywords

big data; data; standardization; national standard; international standard; ISO; reference architecture; data quality; artificial intelligence

References

1. ISO/IEC 20546:2019 Information technology – Big data – Overview and vocabulary) // International Organization for Standardization. URL: <https://www.iso.org/standard/68305.html> (accessed: 01.10.2021).
2. GOST R ISO/MEK 20546–2021 «Informacionnye tekhnologii. Bol'shie dannye. Obzor i slovar'» // Rossijskij institut standartizacii. URL: <https://www.gostinfo.ru/catalog/Details/?id=6859575> (accessed: 01.10.2021).
3. ISO/IEC DIS 22989 Information technology – Artificial intelligence – Artificial intelligence concepts and terminology // International Organization for Standardization. URL: <https://www.iso.org/standard/74296.html> (accessed: 01.10.2021).
4. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh Ch., Byers A.H. Big data: The next frontier for innovation, competition, and productivity // McKinsey Global Institute. P. 12. URL: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#> (accessed: 01.10.2021).
5. Big Data interoperability Framework. V1.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V1_output_docs.php (accessed: 01.10.2021).
6. Big Data interoperability Framework. V2.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V2_output_docs.php (accessed: 01.10.2021).
7. Big Data interoperability Framework. V3.0 Final Version // National Institute of Standards and Technology. URL: https://bigdatawg.nist.gov/V3_output_docs.php (accessed: 01.10.2021).
8. Big Data: Big today, normal tomorrow. ITU T Technology Watch Report. November 2013. // International Telecommunication Union. URL: https://www.itu.int/en/ITU-T/techwatch/Pages/big_data_standards.aspx (accessed: 01.10.2021).
9. Big data: Preliminary Report 2014. ISO/IEC JTC1, 2015. // International Organization for Standardization. URL: https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/big_data_report-jtc1.pdf (accessed: 01.10.2021).
10. ITU-T Y.3600 (11/2015) Big data – Cloud computing based requirements and capabilities. // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=12584&lang=en> (accessed: 01.10.2021).
11. ITU-T Y Suppl. 40 (07/2016) Big data standardization roadmap // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=13022&lang=en> (accessed: 01.10.2021).
12. ISO/IEC TR 20547-2:2018 Information technology – Big data reference architecture – Part 2: Use cases and derived requirements // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-2:ed-1:v1:en> (accessed: 01.10.2021).
13. ISO/IEC TR 20547-5:2018 Information technology – Big data reference architecture – Part 5: Standards roadmap // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-5:ed-1:v1:en> (accessed: 01.10.2021).
14. ISO/IEC TR 20547-1:2020 Information technology – Big data reference architecture – Part 1: Framework and application process // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:20547:-1:ed-1:v1:en> (accessed: 01.10.2021).
15. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20547:-3:ed-1:v1:en> (accessed: 01.10.2021).
16. ISO/IEC 20547-4:2020 Information technology – Big data reference architecture – Part 4: Security and privacy // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20547:-4:ed-1:v1:en> (accessed: 01.10.2021).

17. Walshe R. The Road to Big Data Standardisation // The Elements of Big Data Value / E. Curry et al. (eds.). Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-68176-0_14.
18. ISO/IEC WD 5259-1 Data quality for analytics and ML – Part 1: Overview, terminology, and examples // International Organization for Standardization. URL: <https://www.iso.org/standard/81088.html> (accessed: 01.10.2021).
19. ISO/IEC AWI 5259-2 Data quality for analytics and ML – Part 2: Data quality measures // International Organization for Standardization. URL: <https://www.iso.org/standard/81860.html> (accessed: 01.10.2021).
20. ISO/IEC WD 5259-3 Data quality for analytics and ML – Part 3: Data quality management requirements and guidelines // International Organization for Standardization. URL: <https://www.iso.org/standard/81092.html> (accessed: 01.10.2021).
21. ISO/IEC WD 5259-4 Data quality for analytics and ML – Part 4: Data quality process framework // International Organization for Standardization. URL: <https://www.iso.org/standard/81093.html> (accessed: 01.10.2021).
22. ITU-T Y.3603 Big data – Requirements and conceptual model of metadata for data catalogue» // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14137&lang=en> (accessed: 01.10.2021).
23. ITU-T Y.3604 Big data – Overview and requirements for data preservation» // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14138&lang=en> (accessed: 01.10.2021).
24. Resolution 166. Registration of a PWI entitled «Information technology – Artificial intelligence – Data life cycle framework». ISO/IEC JTC 1/SC 42 “Artificial intelligence”. Resolutions taken during the Closing Plenary - JTC 1/SC 42 - 7 May 2021.
25. ITU-T Y.3651 Big-data-driven networking – mobile network traffic management and planning // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=13818&lang=en> (accessed: 01.10.2021).
26. ITU-T Y.3653 Big data driven networking – functional architecture // International Telecommunication Union. URL: <https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14615&lang=en> (accessed: 01.10.2021).
27. Prikaz Federal'nogo agentstva po tekhnicheskomu regulirovaniyu i metrologii (Rosstandart) ot 25.07.2019 № 1732 (red. ot 20.01.2021) «O sozdanii tekhnicheskogo komiteta po standartizacii “Iskusstvennyj intellekt”» // Rosstandart. <https://www.rst.gov.ru/portal/gost/home/activity/documents/orders#/order/104460> (accessed: 01.10.2021).
28. Prikaz Federal'nogo agentstva po tekhnicheskomu regulirovaniyu i metrologii (Rosstandart) ot 20.08.2020 № 1415 «O vnesenii izmenenij v prikaz Federal'nogo agentstva po tekhnicheskomu regulirovaniyu i metrologii ot 25 iyulya 2019 g. № 1732 «O sozdanii tekhnicheskogo komiteta po standartizacii “Iskusstvennyj intellekt”» // Rosstandart. <https://www.gost.ru/portal/gost/home/activity/documents/orders#/order/179415> (accessed: 01.10.2021).
29. Rasporyazhenie Pravitel'stva RF ot 23.03.2018 N 482-r (red. ot 28.05.2020) «Ob utverzhdenii plana meropriyatij («dorozhnoj karty») po sovershenstvovaniyu zakonodatel'stva i ustraneniyu administrativnyh bar'erov v celyah obespecheniya realizacii Nacional'noj tekhnologicheskoy iniciativy po napravleniyu «Tekhnet» (peredovye proizvodstvennye tekhnologii)» // SZ RF. 2018. № 15 (ch. V). St. 2173.
30. Rossijskaya venchurnaya kompaniya. https://www.rvc.ru/upload/iblock/ac9/Long-term_plan_of_standardization_Teknet.pdf (accessed: 01.10.2021).
31. Prikaz Rosstandarta ot 05.02.2019 g. № 166 «O vnesenii izmenenij v Programmu nacional'noj standartizacii na 2019 god, utverzhdyonnuyu prikazom Federal'nogo agentstva po tekhnicheskomu regulirovaniyu i metrologii ot 1 noyabrya 2018 g. № 2285» // Rosstandart. URL: <https://www.gost.ru/portal/gost/home/activity/standardization> (accessed: 01.10.2021).
32. Protokol rezul'tatov obshchestvennogo obsuzhdeniya effektivnosti primeneniya prinyatyh vo ispolnenie planov meropriyatij «dorozhnyh kart» po sovershenstvovaniyu zakonodatel'stva i ustraneniyu administrativnyh bar'erov v celyah obespecheniya realizacii Nacional'noj tekhnologicheskoy iniciativy normativnyh pravovyh aktov i dokumentov po standartizacii, dostignutyh celej ih primeneniya ne menee chem v techenie odnogo goda s daty nachala

- применения (realizacii) akta // <https://nti2035.ru/upload/351708.2.protocol.pdf> (accessed: 01.10.2021).
33. Dorozhnaya karta razvitiya «skvoznoj» cifrovoj tekhnologii «Nejrotekhnologii i iskusstvennyj intellekt» // Ministerstvo cifrovogo razvitiya, svyazi i massovyh kommunikacij Rossijskoj Federacii. URL: <https://digital.gov.ru/ru/documents/6658/> (accessed: 01.10.2021).
 34. Pasport Federal'nogo proekta «Normativnoe regulirovanie cifrovoj sredy» // Ministerstvo ekonomicheskogo razvitiya Rossijskoj Federacii. URL: https://www.economy.gov.ru/material/directions/gosudarstvennoe_upravlenie/normativnoe_regulirovanie_cifrovoy_sredy/ (accessed: 01.10.2021).
 35. Nacional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda, utverzhdyonnaya Ukazom Prezidenta RF ot 10.10.2019 № 490 «O razvitii iskusstvennogo intellekta v Rossijskoj Federacii» // SZ RF. 2019. № 41. St. 5700.
 36. Federal'nyj proekt «Iskusstvennyj intellekt» // Ministerstvo ekonomicheskogo razvitiya Rossijskoj Federacii. URL: https://www.economy.gov.ru/material/directions/tehnologicheskoe_razvitie/federalnyy_proekt_iskusstvennyy_intellekt/ (accessed: 01.10.2021).
 37. Perspektivnaya programma standartizacii po prioritetnomu napravleniyu «iskusstvennyj intellekt» na period 2021-2024 gody, utverzhdyonnaya Ministerstvom ekonomicheskogo razvitiya Rossijskoj Federacii i Federal'nym agentstvom po tekhnicheskomu regulirovaniyu i metrologii 22.12.2020 g. // URL: https://www.economy.gov.ru/material/news/v_rossii_poyavyatsya_standarty_v_oblasti_iskusstvennogo_intellekta.html (accessed: 01.10.2021).
 38. Okonchatel'naya redakciya proekta GOST R «Informacionnye tekhnologii. Etalonnaya arhitektura bol'shih dannyh. CHast' 1. Struktura i process primeneniya» // Tekhnicheskij komitet 164 «Iskusstvennyj intellekt». URL: <https://www.tc164.ru/окончательные-редакции>. (accessed: 01.10.2021).
 39. Okonchatel'naya redakciya proekta GOST-R «Informacionnye tekhnologii. Etalonnaya arhitektura bol'shih dannyh. CHast' 2. Varianty ispol'zovaniya i proizvodnye trebovaniya» // Tekhnicheskij komitet 164 «Iskusstvennyj intellekt». URL: <https://www.tc164.ru/окончательные-редакции> (accessed: 01.10.2021).
 40. Proekt PNST «Informacionnye tekhnologii. Bol'shie dannye. Tipovaya arhitektura» // Elektronnyj fond pravovyh i normativno-tekhnicheskikh dokumentov. URL: <https://docs.cntd.ru/document/554715621> (accessed: 01.10.2021).
 41. Pervaya redakciya proekta GOST R «Informacionnye tekhnologii. Etalonnaya arhitektura bol'shih dannyh. CHast' 5. Napravleniya standartizacii» // Tekhnicheskie komitet 164 «Iskusstvennyj intellekt». URL: <https://www.tc164.ru/первые-редакции> (accessed: 01.10.2021).
 42. Okonchatel'naya redakciya proekta GOST R «Informacionnye tekhnologii. Bol'shie dannye. Tekhnicheskoe zadanie. Trebovaniya k sodержaniyu i oformleniyu» // Tekhnicheskij komitet 164 «Iskusstvennyj intellekt». URL: <https://www.tc164.ru/окончательные-редакции> (accessed: 01.10.2021).
 43. Pervaya redakciya proekta GOST R «Informacionnye tekhnologii – Iskusstvennyj intellekt – Struktura upravleniya processami analitiki bol'shih dannyh» // Tekhnicheskij komitet 164 «Iskusstvennyj intellekt». URL: <https://www.tc164.ru/первые-редакции> (accessed: 01.10.2021).
 44. ISO/IEC 24668 Information technology – Artificial intelligence – Process management framework for big data analytics // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:24668:dis:ed-1:v1:en> (accessed: 01.10.2021).
 45. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis // Scientometrics. 2020. Vol 122. P. 1563–1581. <https://doi.org/10.1007/s11192-020-03371-2>.
 46. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologij raboty s bol'shimi dannymi // Informacionnoe obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
 47. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends // J Big Data. 2017. Vol 4, № 30. <https://doi.org/10.1186/s40537-017-0088-1>.
 48. de Meer, J. A Theory on Big Data. // INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge) / Draude, C., Lange, M. & Sick, B. (Hrsg.). Bonn: Gesellschaft für Informatik e.V. 2019. S. 261-269. https://doi.org/10.18420/inf2019_ws30.

49. Walshe R., Casey K., Kernan J., Fitzpatrick D. AI and Big Data Standardization: Contributing to United Nations Sustainable Development Goals // Journal of ICT Standardization: 2020: Vol 8 Iss 2. URL: <https://journals.riverpublishers.com/index.php/IJCTS/article/view/2649> (accessed: 01.10.2021).
50. Caballero I., Parody L., Bermejo I., Gómez López M.T., Gasca R.M., Piattini M. Service level agreement for data quality governed by Iso 8000-1X0 // Proceedings of the 19th International Conference on Information Quality, ICIQ. 2014. P. 114–127.
51. Xiao Y., Lu L.Y.Y., Liu J.S., Zhou Z. Knowledge diffusion path analysis of data quality literature: A main path analysis // Journal of Informetrics. 2014. Vol 8 Iss 3. P. 594–605. <https://doi.org/10.1016/j.joi.2014.05.001>.
52. Perin U. Reference Architectures and Standards for the Internet of Things and Big Data in Smart Manufacturing // International Journal of Recent Technology and Engineering. 2019. Vol 8 Iss 1C2. P. 884–887. <https://doi.org/10.1109/FiCloud.2019.00041>.
53. Heale B.S.E., Overby C.L., Del Fiol G. et al. Integrating genomic resources with electronic health records using the HL7 infobutton standard // Applied Clinical Informatics. 2016. Vol 7 Iss 3. P. 817–831. <https://doi.org/10.4338/ACI-2016-04-RA-0058>.
54. Raghunath N. A Standard for Benchmarking Big Data Systems // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. Vol 8585. P. 193–201. https://doi.org/10.1007/978-3-319-10596-3_15.
55. Laney, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Research Note, 6, February 2001. URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed: 01.10.2021).
56. Cloutier R., Muller G., Verma D., Nilchiani R., Hole E., Bone M. The Concept of Reference Architecture // Systems Engineering. 2009. Vol 13, Iss 1. P 14–27. <https://doi.org/10.1002/sys.20129>.
57. Big data: The next frontier for innovation, competition, and productivity. Report // McKinsey Global Institute. 2011. May. URL: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (accessed: 01.10.2021).
58. Mohd Rehan Ghazi Durgaprasad Gangodkar Hadoop, MapReduce and HDFS: A Developers Perspective // Procedia Computer Science. 2015. Vol 48. P. 45–50. <https://doi.org/10.1016/j.procs.2015.04.108>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050915006171> (accessed: 01.10.2021).
59. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters // OSDI'04: Sixth Symposium on Operating System Design and Implementation. San Francisco, CA, 2004. P. 137–150. URL: <https://static.googleusercontent.com/media/research.google.com/ru//archive/mapreduce-osdi04.pdf> (accessed: 01.10.2021).
60. GOST R ISO/MEK 27002–2021 Informacionnye tekhnologii. Metody i sredstva obespecheniya bezopasnosti. Svod norm i pravil primeneniya mer obespecheniya informacionnoj bezopasnosti // Elektronnyj fond pravovyh i normativno-tekhnicheskikh dokumentov. URL: <https://docs.cntd.ru/document/1200179669> (accessed: 01.10.2021).
61. GOST R 56214–2014/ISO/TS 8000-1:2011 Kachestvo dannyh. CHast' 1. Obzor // Elektronnyj fond pravovyh i normativno-tekhnicheskikh dokumentov. URL: <https://docs.cntd.ru/document/1200114769?section=text> (accessed: 01.10.2021).
62. ISO/IEC 25012:2008 Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:25012:ed-1:v1:en> (accessed: 01.10.2021).
63. ISO 9000:2015 Quality management systems – Fundamentals and vocabulary // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:ru> (accessed: 01.10.2021).
64. Pirsig R. Dzen i iskusstvo uhoda za motociklom/ per. s angl. M.Gorshkova. – SPb.: «Simpozium», 2002.
65. David J. Hand. Dark Data. Princeton University Press. Princeton and Oxford., 2020.
66. Dirty Data // Techopedia. URL: <https://www.techopedia.com/definition/1194/dirty-data> (accessed: 01.10.2021).

67. Raikov A.N., Avdeeva Z., and Ermakov A. Big Data Refining on the Base of Cognitive Modeling // Proceedings of the 1st IFAC Conference on Cyber-Physical&Human-Systems, Florianopolis, Brazil. 2016. P. 147-152. <https://doi.org/10.1016/j.ifacol.2016.12.205>.
68. Raikov A. Cognitive Semantics of Artificial Intelligence: A New Perspective // Springer Singapore, Topics: Computational Intelligence XVII, 2021. <https://doi.org/10.1007/978-981-33-6750-0>.
69. Rajkov A.N., Kotel'nikov V.A. Predvoskhishchenie budushchego cifrovizacii // Informatizaciya i svyaz'. 2019. № 1. – С. 7 – 11. <https://doi.org/10.34219/2078-8320-2019-10-1-7-11>.
70. ISO/IEC 2382:2015 Information technology – Vocabulary // International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en> (accessed: 01.10.2021).

Технологии информационного общества

МОДЕЛЬ BD4DE-ММ ЗРЕЛОСТИ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ В ОРГАНИЗАЦИИ

Ершов Петр Сергеевич

*Институт развития информационного общества, руководитель Дирекции региональных программ
Москва, Российская Федерация
peter.erшов@iis.ru*

Катин Александр Владимирович

*Институт развития информационного общества, руководитель Дирекции отраслевых программ
РЭУ имени Г. В. Плеханова, старший преподаватель Базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
alexander.katin@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель Совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель Базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, Лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Аннотация

Предложена модель зрелости BD4DE-ММ использования технологий работы с большими данными в организации с акцентом на достижение социальных и экономических эффектов. Проведен анализ существующих моделей зрелости использования цифровых технологий в организации, выделены основные размерности: стратегия и регулирование; кадры и лидерство; данные; инструменты и аналитика; инфраструктура и безопасность; организация работы; воздействие. Предложены индикаторы, позволяющие по каждой размерности оценить по 5-балльной шкале уровень зрелости работы с большими данными в организации.

Ключевые слова

модель зрелости; большие данные; ключевые факторы успеха; использование больших данных; аналитика больших данных; BD4DE-ММ; Big Data for Digital Economy Maturity Model; стратегия; регулирование; кадры; лидерство; данные; инструменты; инфраструктура; безопасность; организация работы; воздействие

Введение

Современные процессы социально-экономического развития, основанные на масштабном использовании цифровых технологий, радикально трансформируют эти процессы, причем скорость изменений возрастает. Многие страны связывают надежды на достижение социальных и экономических эффектов с процессами цифровой трансформации экономики, основанными на применении цифровых технологий [третьей волны] – таких как технологии искусственного

© Ершов П.С., Катин А.В., Хохлов Ю.Е. Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_259

интеллекта, предиктивной аналитики, распределенного реестра, робототехники или интернета вещей [1].

Исследования последних лет показали, что развитие цифровой экономики и извлечение цифровых дивидендов [2] – социальных и экономических эффектов от использования цифровых технологий – невозможно без учета нецифровых факторов, таких как государственная политика и стратегическое планирование, нормативное регулирование, развитие человеческого капитала, исследования и разработки в области цифровых технологий. В 2017 году Всемирным банком в сотрудничестве с Институтом развития информационного общества (ИРИО) и другими партнерами в России была разработана методика оценки уровня развития цифровой экономики в стране (Digital Economy Country Assessment, DECA) и апробирована на примере Российской Федерации [3]. Методология DECA построена на анализе результатов исследований, полученных международными организациями (Всемирным банком, ОЭСР, МСЭ, Всемирным экономическим форумом), ведущими мировыми консалтинговыми компаниями, представителями отраслей и нашла дальнейшее применение как на уровне отдельной страны, так и на уровне региона или города [4].

В цифровую эпоху эффективность использования цифровых технологий в отдельно взятой организации существенно влияет как на всю отрасль экономики или сферу деятельности, так и на страну в целом. Поэтому представляется важным оценить уровень проникновения/освоения тех или иных цифровых технологий в организации и эффекты, которые они приносят самой организации и потребителям ее продуктов и услуг.

Одним из распространенных инструментов оценки уровня использования ИКТ в целом (и технологий работы с большими данными в частности) в последние десятилетия стали модели зрелости, позволяющие выявить сильные и слабые стороны проникновения и освоения технологических инноваций в организации, оценить текущее состояние и определить направления развития.

Построение разнообразных моделей зрелости, как правило, восходит к подходу, первоначально примененному в Университете Карнеги-Меллона (Carnegie Mellon University) для оценки процессов разработки программного обеспечения и со второй половины 1980-х годов развиваемому под эгидой Software Engineering Institute (SEI). Первоначальная модель зрелости процессов разработки ПО – Capability Maturity Model for Software (CMM) – была впоследствии расширена до методологии (модели) совершенствования процессов в организациях разных размеров и видов деятельности, которая получила название Capability Maturity Model Integration (CMMI) [5]. Поддержка в актуальном состоянии и развитие методологии CMMI осуществлялась некоммерческой организацией CMMI Institute [6] при университете Карнеги-Меллона, которая с 2016 года стала подразделением коммерческой ассоциации ISACA.

Полезность и гибкость подхода, связанного с оценкой уровня зрелости, сделало популярным использование подобных моделей в других сферах деятельности и привело к их стандартизации на международном уровне [7, 8]. В Российской Федерации также действует идентичный национальный стандарт, фиксирующий требования к эталонным моделям процесса, моделям оценки процесса и моделям зрелости [9, 10]. Под *моделью зрелости* понимается модель, извлеченная из одной или более указанных моделей оценки процесса, которая определяет составы процессов, связанные с определенными уровнями организационной зрелости процесса. В общем случае модель зрелости в той или иной сфере деятельности включает определенные области оценки, которые называются *размерностями* (dimensions) или *ключевыми факторами успеха* (key success factors), в которых могут выделяться отдельные аспекты (aspects) и характеризующие их *индикаторы оценки* (assessment indicators), а также *шкалу организационной зрелости* от самого низкого, базового уровня до самого высокого. Под *уровнем зрелости* будет пониматься состояние, в котором достигнута определенная организационная цель и не требуется каких-либо дальнейших действий в этом направлении. Но это не означает, что организации нужно остановиться в своем развитии – для достижения новой цели, возможно, потребуется следующий уровень зрелости, который использованная ранее модель не учитывала.

В данной работе исследование проводится применительно к цифровым технологиям работы с большими данными, значимость которых в цифровую эпоху стала общепризнанной [11]. Основная цель – построить модель зрелости использования цифровых технологий работы с большими данными в организации для достижения социальных и экономических эффектов с учетом концептуальных положений методологии DECA.

1 Выделение ключевых факторов успеха использования технологий работы с большими данными в организации

1.1 Обзор литературы по моделям зрелости использования цифровых технологий в организациях

Построение модели зрелости начинается с выделения совокупности ключевых факторов, влияющих на успех использования всего стека цифровых технологий в организациях в целом и технологий работы с большими данными в частности. Отбор потенциальных факторов успеха проводился в ходе анализа релевантных научных публикаций и открытых аналитических материалов. В качестве источников были использованы международные базы данных индексируемых научных публикаций Scopus, Web of Science (WoS), а также открытые базы публикаций www.researchgate.net; academia.org и Google Scholar.

Для отбора базового массива публикаций был сформирован поисковый образ, состоящий из логического объединения двух поисковых запросов: «цифровые технологии» и «модель зрелости».

Поисковый запрос «цифровые технологии» включал следующие ключевые слова и словосочетания: digital technology; big data; big data analytics; industry 4.0; artificial intelligence; machine learning; deep learning; blockchain; distributed ledger technology; robotics; internet of things; virtual reality; augmented reality, additive manufacturing; 5G; mobile technology.

В поисковый запрос «модель зрелости» вошли следующие ключевые слова и словосочетания: maturity model; maturity scale; maturity level; maturity framework; critical success factor; digital transformation; digital maturity, key factor.

Поисковый образ формировался логическим объединением двух поисковых запросов с помощью логического оператора «AND», а в рамках одного поискового запроса ключевые слова и словосочетания использовался логический оператор «OR».

Применение поискового образа для отбора базового массива публикаций из перечисленных выше источников привело к большому массиву: более тысячи публикаций по каждому источнику со значительным пересечением по результатам. Поэтому для дальнейшего анализа научных публикаций в качестве основного источника была выбрана социальная сеть и платформа для сотрудничества научных работников www.researchgate.net, содержащая, как правило, не только метаданные с библиографическими описаниями публикаций, но и полные тексты. Аналитические материалы и конкретные модели зрелости, разработанные консалтинговыми компаниями, за редким исключением находятся в открытом доступе во Всемирной паутине и были найдены в результате онлайн-поиска в Google Scholar.

В результате применения поискового образа к базе данных www.researchgate.net был получен массив из 1325 публикаций. В то же время поиск в Google Scholar позволил дополнить массив 418 релевантными документами: моделями зрелости работы с данными, моделями зрелости работы с большими данными и аналитическими материалами консалтинговых компаний и исследовательских организаций.

Дальнейший поиск наиболее релевантных документов из полученного массива 1743 публикаций состоял в применении уточняющих запросов. Так, для уточненного поискового запроса «цифровые технологии» были выбраны ключевые слова: big data; big data analytics; industry 4.0; internet of things, artificial intelligence, а уточненный поисковый запрос «модель зрелости» был ограничен набором ключевых слов: maturity model, maturity scale, success factor, key factor.

Применение уточненных поисковых запросов к ранее полученной выборке снизило число отобранных документов до 221, после чего были проанализированы все аннотации на предмет соответствия целям настоящего исследования, прежде всего, с точки зрения разработки и/или изучения моделей зрелости различных процессов. По результатам анализа аннотаций были отобраны 63 документа, полные тексты которых также были подвергнуты анализу. По итогам описанной выше многоступенчатой процедуры были получены следующие результаты: в базе данных www.researchgate.net нашлись 17 публикаций [12-28], посвященных оценке уровня зрелости использования цифровых технологий в организациях; а в Google Scholar были найдены 5 моделей зрелости работы с данными [29-33]; 4 модели зрелости работы с большими данными [34-37], одна модель зрелости использования технологий искусственного интеллекта [38] и 5 аналитических материалов, касающихся уровня зрелости использования различных цифровых технологий в организациях [39-43].

Для дальнейшего анализа отобранные 32 документа были разделены на три категории: модели зрелости и концептуальные подходы к оценке уровня зрелости использования цифровых технологий в организациях (6 документов); модели зрелости управления данными в организациях (6 документов); модели зрелости работы с большими данными в организациях (20 документов). Выделение ключевых факторов успеха, влияющих на использование больших данных в организации, проводился для каждой категории документов отдельно.

1.2 Зрелость использования цифровых технологий в организации

Первая категория отобранных на предыдущем этапе документов связана с построением моделей зрелости использования любых цифровых технологий (не обязательно для работы с большими данными) в организации. Документы были проанализированы на предмет выделения размерностей (факторов) и отдельных аспектов, влияющих на уровень зрелости. Выделенные размерности и/или отдельные аспекты в каждом документе представлены ниже в таблице 1.

Таблица 1. Модели зрелости использования цифровых технологий в организации

№	Модель зрелости	Размерности / Аспекты
1.	Индекс зрелости Индустрии 4.0 [29]	Материальные и нематериальные ресурсы (сотрудники компании; станки и оборудование; инструменты; материалы; конечный продукт); Информационные системы (подготовка и обработка данных, пригодных для принятия решений; интеграция для более эффективного использования данных и повышения гибкости их обработки); Организационная структура (органическая внутренняя организация; динамичное сотрудничество в цепочке создания ценности); Культура организации (готовность к изменениям; социальное взаимодействие); Функциональные области в бизнесе (разработка; производство; логистика; обслуживание; маркетинг и продажи)
2.	A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises [27]	Стратегия; Лидерство; Клиенты; Продукты; Операции; Культура; Люди; Управление; Технологии
3.	Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises [28]	Технологии; Продукты; Клиенты и Партнеры; Процесс создания добавленной стоимости; Информация и данные; Корпоративные стандарты; Сотрудники; Стратегия и лидерство
4.	Industry 4.0 Readiness Online Self-Check for Businesses [30]	Стратегия; Смарт-фабрика; Смарт-операции; Ориентация на данные; Сотрудники
5.	TM Forum. AI Maturity Model Toolkit [38]	Стратегия (портфолио и идеи, финансы и инвестиции, стратегическое управление); Культура, (корпоративная культура, лидерство, организационная структура и управление талантами); Пользователи (вовлеченность, опыт, доверие); Операции (управление жизненным циклом услуг, гарантия обслуживания с обратной связью, управление процессами); Данные (управление работой с данными, курирование данных, капитализация данных); Технологии (развитие и внедрение, архитектура ИИ-систем, автоматизация и приложения)
6.	TDWI IoT Data Readiness Guide [40]	Организация (варианты использования в бизнесе; лидерство и культура; стратегия; команда); Данные (масштаб; разнообразие; частота; интеграция; управление); Инфраструктура (масштабируемость; технологии обработки; технологии хранения; безопасность и доверие); Аналитика (аналитика в реальном времени; продвинутая аналитика; встроенная аналитика; развертывание аналитики); Информационные технологии, разработка и внедрение (изменение бизнес-процессов; инструменты; навыки)

Проведенный анализ показал, что в рассмотренных моделях превалирует использование размерностей «Стратегия», «Культура (организации)», «Кадры» и «Технологии». Для оценки организационной зрелости при использовании цифровых технологий традиционно используются 5-балльные шкалы, хотя сами названия уровней зрелости, да и количество баллов для отнесения к тому или иному уровню могут сильно варьироваться.

1.3 Зрелость управления данными в организации

При анализе второй группы отобранных публикаций, содержащих описание конкретных моделей зрелости работы с данными в организациях, также были выделены основные факторы и аспекты, приведенные ниже в таблице 2.

Таблица 2. Модели зрелости управления данными в организации

№	Модель зрелости	Размерности / Аспекты
1	Federal Government Data Maturity Model [47]	Возможности аналитики данных; Культура обращения с данными; механизмы управления данными; Персонал для управления данными; Системы/технологии управления данными; Организационная структура, отвечающая за управление данными
2	Модель зрелости управления данными MM-SERV [48]	Доступность данных; Использование данных; Состояние данных; Доступность сервисов; Удобство сервисов; Мониторинг данных
3	TDWI Analytics Maturity Model Assessment [39]	Организация; Ресурсы; Инфраструктура данных; Аналитика; Управление работой с данными
4	Gartner's Enterprise Information Management Maturity Model [12, 33]	Видение; Стратегия; Метрики; Управление; Организация; Процессы; Технологии
5	TDWI Self-Service Analytics Maturity Model Guide [43]	Организация (лидерство; стратегия; достижения; воздействие; культура); Менеджмент данных (доступ; предобработка; интеграция; качество); Инфраструктура (команда; технологии; архитектура; производительность); Аналитика (масштаб; доставка; обучение; навыки); Управление (политика; структура; регламентирование; координация; безопасность и доверие)
6	Data Maturity Framework for the Not-For-Profit Sector [32]	Использование (цели сбора и анализ; преимущества и эффекты); Данные (сбор; качество; источники; активы); Анализ (тип; техника; присоединение; представление); Лидерство (отношение; планы; возможность; инвестиции); Культура (командный подход; самообследование; открытость; защита); Инструменты (место хранения; тип; качество; обмен; интеграция); Навыки (возможности; навыки и умения; обучение; доступ к знаниям и экспертиза)

В моделях зрелости, представленных в этой группе, под прицелом находится оценка организационной зрелости управления данными. Как и в предыдущем разделе, в качестве основных размерностей выделяются: «Стратегия» и «Кадры», но при этом акцент делается на новые размерности «Данные», «Организация работы» и «Инструменты (включая аналитику)». В качестве шкал организационной зрелости при использовании цифровых технологий также используются 5-балльные шкалы.

1.4 Зрелость работы с большими данными в организации

Наиболее релевантными для данного исследования являются разработанные ранее модели зрелости работы с большими данными, подавляющее большинство которых было

проанализировано и отражено ниже в таблице 3. Там же представлены размерности и аспекты, содержащиеся в анализируемых моделях зрелости

Таблица 3 – Модели зрелости использования больших данных в организации

№	Модель зрелости	Размерности / Аспекты
1	IBM Big Data & Analytics Maturity Model [37]	Стратегия; Информация и данные; Аналитика; Человеческий капитал; Управление
2	A Guide to Achieving Big Data Analytics Maturity [34]	Институты управления и финансирования; Цифровая инфраструктура; Человеческий капитал; Политика и регулирование; Аналитика
3	Hortonworks Big Data Maturity Model [35]	Политика и регулирование; Управление и финансирование; Человеческий капитал; Информационная безопасность; Цифровая инфраструктура
4	IDC MaturityScape Benchmark: Big Data and Analytics in the United States [36]	Стратегия; Данные; Технологии; Человеческий капитал; Процессы
5	TDWI Advanced Analytics Maturity Model Guide [41]	Организация (лидерство; культура; эффекты; стратегия); Ресурсы (финансирование; таланты и навыки; роли и ответственность; обучение); Инфраструктура данных (разнообразие, объем, скорость; доступ к данным; интеграция данных; архитектура данных); Аналитика (Варианты использования; техники; подходы к развертыванию и доставке; управление); Управление (организационная структура; политика процессы и контроль исполнения; безопасность)
6	TDWI Hadoop Readiness Guide [42]	Организация (лидерство; стратегия; текущее состояние; корпоративное руководство; конкретные варианты использования; потенциальные варианты использования); Большие данные (наличие больших данных; объем больших данных; аналитика как путь к ценности; сбор больших данных; управление объемами данных); Управление большими данными (команда управляющая данными; навыки специалистов в области больших данных; владение и управление данными; управление объемами данных; инфраструктура интеграции данных; богатство источников данных; богатство структур данных); Аналитика (команда аналитики; приверженность аналитике; операционная аналитика; компетенции аналитиков; компетенции программистов; компетенции в области распределённых вычислений; разнообразие аналитики); Информационные технологии (соответствие информационных технологий; соответствие команды имеющимся ИТ-решениям; ИТ-инновации; информационная безопасность; ИТ-экспертиза и навыки)
7	Critical Success Factor Categories for Big Data: A Preliminary Analysis of the Current Academic Landscape [13]	Организация; Люди; Технологии
8	Critical Success Factors of Business Intelligence and Big Data Analysis: Utilizing Hidden Business Value	Инфраструктура; Организация; Производительность; Эффекты

№	Модель зрелости	Размерности / Аспекты
	from Big Data for Business Intelligence [14]	
9	An Investigation into the Implementation Factors Affecting the Success of Big Data Systems [15]	Управление данными; Доверие и безопасность; Управление предприятием; Организация; Культура; Масштабирование; Навыки; Технологии
10	Amazon in the Air: Innovating with Big Data at Lufthansa [16]	Эффекты; Лидерство; Сервис-ориентированный подход; Отказ от модных ИТ-тенденций; Интеграция; Управление компетенциями
11	Big Data Project Success – A Meta Analysis [17]	Стратегия; Эффекты; Лидерство; Навыки; Качество данных, безопасность и интеграция; Технологические возможности
12	Big Data Team Process Methodologies: A Literature Review and the Identification of Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project's Success [18]	Данные; Управление; Процессы; Цели; Команда; Инструменты
13	Factors Influencing to the Implementation Success of Big Data Analytics: A Systematic Literature Review [19]	Организация; Технологии; Люди; Аналитика; Культура; Среда; Управление данными; Качество данных; Приложения; Эффекты
14	Towards a Process View on Critical Success Factors in Big Data Analytics Projects [20]	Люди; Процессы; Технологии
15	Critical success factors for Big Data adoption in the virtual retail: Magazine Luiza case study [21]	Стратегическое выравнивание; Вовлеченность высшего руководства; Организационная структура; Обмен информацией; Культурные изменения в сторону ориентации на данные и эксперименты; Наличие квалифицированных специалистов; Коммуникация; Ожидания руководства; Гибкие методы управления; Доверие и безопасность
16	Conceptual Model for Successful Implementation of Big Data in Organizations [22]	Поддержка высшего руководства; Организационные изменения; Доступность и качество данных; Инфраструктура; Навыки; Безопасность и доверие
17	Measuring the Barriers to Big Data for Development: Design-Reality Gap Analysis in Colombia's Public Sector [23]	Информация; Технологии; Процессы; Цели и ценности; Навыки и знания; Система управления
18	Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare [24]	Данные; Технологии; Организация; Поддержка

№	Модель зрелости	Размерности / Аспекты
19	On the Adoption of Big Data Analytics: Interdependencies of Contextual Factors [25]	Технологии; Организация; Среда
20	Big Data analytics: Understanding its capabilities and potential benefits for healthcare organizations [26]	Инфраструктура; Бизнес-процессы; Организация; Управление; Стратегия; Эффекты

В этой группе собрано большинство моделей зрелости работы с большими данными, наиболее релевантные предмету настоящего исследования. Наряду с отмеченными выше в разделах 1.2–1.3 размерностями «Стратегия», «Кадры», «Организация работы» в анализируемых моделях особое внимание уделяется размерностям «Данные», «Технологии», «Инструменты» и «Аналитика», которые адаптированы, прежде всего, к технологиям работы с большими данными. Что касается шкал зрелости, то, следуя традиции, многие модели придерживаются 5-балльных порядковых шкал, берущих свое начало в первых моделях зрелости [5].

1.5 Семантический и частотный анализ моделей зрелости

Приведенные в разделах 1.2–1.4 результаты анализа используемых в моделях зрелости размерностей (ключевых факторов) и отдельных аспектов, влияющих на использование цифровых технологий в организациях, с одной стороны демонстрируют присутствие сходных характеристик в различных моделях зрелости, а с другой – терминологическое разнообразие при их описании. Например, наряду с верхнеуровневой размерностью «Стратегия» из одной модели зрелости [17], в другой модели зрелости [41] эта характеристика уже выступает в качестве отдельного аспекта внутри размерности «Организация».

Для дальнейшего анализа была проведена процедура семантического обобщения размерностей моделей из таблиц 1–3, результаты которого представлены ниже в таблице 4.

Таблица 4. Семантический анализ размерностей и аспектов моделей зрелости

Размерность	Семантический аналог
Аналитика	аналитика в реальном времени, продвинутая аналитика, встроенная аналитика, развертывание аналитики, анализ
Безопасность	доверие, доверие и безопасность, информационная безопасность
Воздействие	эффекты, процесс создания добавленной стоимости
Данные	качество данных, подготовка данных, обработка данных, интеграция данных, доступ к данным, архитектура данных, менеджмент данных, и т.д.
Инфраструктура	масштабируемость, технологии обработки, технологии хранения
Кадры	люди, команда, навыки и знания, управление компетенциями
Культура	культура организации, корпоративная культура, командный подход, самообследование, открытость
Лидерство	вовлеченность высшего руководства, поддержка высшего руководства
Организация работы	организационная структура, управление работой с данными, регламентирование, институты управления и финансирования, процессы и контроль исполнения, команда управляющая данными, управление предприятием, гибкие методы управления
Ресурсы	ресурсы, финансирование, инвестиции
Стратегия	цели использования, цели и ценности, стратегическое выравнивание, политика и регулирование
Технологии	развитие технологий, развертывание и вывод из эксплуатации, инструменты, автоматизация и приложения, информационные системы, системы/технологии управления данными

Последующий анализ исследуемых 32 моделей зрелости показал, что наиболее часто в качестве размерностей используются: аналитика; безопасность; воздействие; данные; инструменты; инфраструктура; кадры; лидерство; организация работы; регулирование; стратегия. Сравнение применяемых размерностей с концептуальной схемой оценки уровня развития цифровой экономики DECA [4] показало, что среди них содержится основной набор факторов, влияющих на социально-экономическое развитие, базирующееся на широкомасштабном использовании цифровых технологий. Это послужило дополнительным аргументом в пользу выбора перечисленных выше характеристик для оценки уровня зрелости работы с большими данными в организации.

2 Модель зрелости работы с большими данными в организации

В настоящем разделе представлена модель зрелости использования технологий работы с большими данными в организации для достижения социальных и экономических эффектов (Big Data for Digital Economy Maturity Model, BD4DE-MM), которая учитывает концептуальные положения методологии DECA для оценки развития цифровой экономики и результаты анализа существующих моделей зрелости из раздела 1.

В распространенных моделях зрелости, как правило, используются порядковые шкалы зрелости, наиболее часто встречается 5-балльная шкала. При построении модели зрелости BD4DE-MM также выбрана 5-балльная шкала как интуитивно понятная отечественным респондентам (в частности, из-за 5-балльной шкалы оценок, используемой в российской системе образования).

При этом каждому из уровней будет соответствовать характеристика зрелости организации с точки зрения работы с большими данными.

1. Начальный уровень – организация не использует большие данные.
2. Формирующийся уровень – организация, начинающая использовать большие данные.
3. Средний уровень – организация, использующая большие данные.
4. Продвинутый уровень – организация, масштабно использующая зрелые технологии работы с большими данными.
5. Высокий уровень – организация-лидер, масштабно использующая большие данные и внедряющая новые технологии работы с большими данными.

Выбор размерностей для модели зрелости BD4DE-MM продиктован необходимостью оценки воздействия технологий работы с большими данными в организации на достижение экономических и социальных эффектов. Проведенный анализ существующих моделей и ранее проведенных исследований (см. раздел 1.5) показал, что основные характеристики, влияющие на успешное применение этих технологий, могут быть сгруппированы в следующие семь размерностей:

1. Стратегия и регулирование.
2. Кадры и лидерство.
3. Данные.
4. Инструменты и аналитика.
5. Инфраструктура и безопасность.
6. Организация работы.
7. Воздействие.

Описание и обоснование выбора предложенных размерностей и соответствующих индикаторов уровня зрелости работы с большими данными в модели BD4DE-MM приведены в следующих разделах.

2.1 Стратегия и регулирование

Цифровая трансформация деятельности организации, основанная на работе с большими данными, предполагает наличие корпоративной политики, включающей явно обозначенную стратегию работы с большими данными для достижения целей организации, разработку и реализацию планов действий по ее реализации, проведение регулярного мониторинга и оценки для обеспечения обратной связи в системе управления организацией.

Анализ существующих моделей зрелости (см. раздел 1) показывает, что содержательная оценка стратегии работы организации с большими данными не может ограничиваться только

определением наличия такой стратегии в виде отдельного документа или раздела цифровой стратегии организации. Зрелая стратегия работы с большими данными как инструмент развития организации должна соответствовать общей стратегии развития организации, предполагать наличие плана ее реализации, быть обеспечена необходимыми ресурсами и ориентирована на технологические инновации в области работы с большими данными.

Не менее важным для эффективной работы с большими данными в организации является соблюдение требований правового регулирования и следование современным стандартам работы с большими данными. Изменения нормативного правового и технического регулирования осуществляются гораздо медленнее, чем происходит развитие цифровых технологий, поэтому в последние годы активно развиваются процессы саморегулирования основных заинтересованных сторон при работе с большими данными. Поэтому участие организации в процессах саморегулирования позволяет опережающим темпом формировать и следовать принципам и процедурам, которые впоследствии превращаются в правовые и/или технические нормы.

С учетом сказанного для размерности «Стратегия и регулирование» в модель зрелости BD4DE-MM были включены следующие индикаторы.

1. Наличие в организации стратегии работы с большими данными.
2. Соответствие стратегии работы с большими данными положениям стратегии развития организации.
3. Наличие в организации плана реализации стратегии работы с большими данными.
4. Наличие в организации необходимых ресурсов (кадровых, финансовых и т.д.) для реализации стратегии работы с большими данными.
5. Наличие в стратегии работы с большими данными мероприятий, ориентированных на эксперименты с перспективными технологиями.
6. Соответствие деятельности организации требованиям нормативного правового регулирования работы с большими данными.
7. Соответствие деятельности организации международным/национальным стандартам работы с большими данными.
8. Соответствие деятельности организации принципам саморегулирования работы с большими данными.

2.2 Кадры и лидерство

Размерность «Кадры и лидерство» объединяет аспекты оценки зрелости, которые в экономических и социологических исследованиях принято называть «человеческим капиталом». Под человеческим капиталом традиционно понимается совокупность приобретенных знаний, навыков, опыта и ценностей, которые влияют на экономическую продуктивность человека. Совокупный человеческий капитал и способность к его воспроизведению в организации является важным фактором ее эффективности и возможности дальнейшего развития.

При построении модели зрелости BD4DE-MM человеческий капитал организации как фактор эффективности работы с большими данными рассматривается с точки зрения осведомленности, наличия знаний и компетенций сотрудников организации (включая руководителей высокого уровня) для работы с большими данными и их использования в профессиональной деятельности. Для комплексной характеристики человеческого капитала организации оценивается: (1) существующий в организации человеческий капитал; (2) действующая система воспроизводства человеческого капитала организации (повышение уровня компетенций кадров для работы с большими данными) и ее соответствие потребностям.

Дополнительно в данной размерности рассматривается еще один важный аспект, смежный с оценкой человеческого капитала, который обычно называют «цифровым лидерством» – наличие в организации «лидеров» в работе с большими данными – руководителей высокого уровня, имеющих необходимые навыки работы с большими данными, способных увлечь за собой и направлять других сотрудников, управлять работой с большими данными.

На основе описанного выше подхода для размерности «Кадры и лидерство» в модели зрелости BD4DE-MM были выбраны следующие индикаторы оценки зрелости.

1. Осведомленность о возможностях и навыки работы с большими данными сотрудников организации.
2. Обеспеченность организации специалистами по работе с большими данными.

3. Потребности организации в системе повышения уровня компетенций кадров для работы с большими данными.
4. Наличие в организации «лидеров» в работе с большими данными среди руководителей высокого уровня.
5. Осведомленность о возможностях и навыки работы с большими данными у руководителей высокого уровня.
6. Использование руководителями высокого уровня аналитики больших данных в своей профессиональной деятельности (при принятии решений).

2.3 Данные

Эффективность деятельности организации в настоящее время все больше зависит от успешности работы с (большими) данными.

Детальный анализ моделей зрелости (см. 1.5) показывает, что обеспеченность организации качественными большими данными, наличие и использование мастер-данных, реализация возможности интеграции данных из разных источников, их совместное использование и, наконец, работа с большими данными на всех стадиях их жизненного цикла являются важными факторами ее успешной деятельности. Эти характеристики и были выбраны в качестве основных для набора индикаторов, характеризующих размерность «Данные» для модели зрелости BD4DE-MM.

1. Обеспеченность деятельности организации большими данными.
2. Обеспеченность деятельности организации качественными большими данными.
3. Обеспеченность деятельности организации мастер-данными.
4. Интеграция больших данных в организации.
5. Совместное использование больших данных в организации.
6. Работа с большими данными в организации на стадиях жизненного цикла.

2.4 Инструменты и аналитика

Анализ большинства существующих моделей зрелости выявил два основных направления использования больших данных в организациях: (1) использование специализированных инструментов для работы с большими данными и (2) использование различных видов аналитики больших данных.

Большие данные отличаются такими характеристиками, как большой объем, разнообразие, скорость обработки и/или вариативность. Работа с большими данными предъявляет особые требования к применяемым технологиям и инструментам, выходящим за рамки возможностей традиционных систем управления базами данных и бизнес-аналитики, поэтому использование специализированных инструментов является важнейшим признаком зрелости работы с большими данными в организации.

Главная цель работы с большими данными – это извлечение необходимых знаний из имеющихся больших массивов данных для повышения эффективности деятельности организации. Поэтому ключевым инструментом при работе с большими данными является активное использование аналитики больших данных в основных, обеспечивающих и управленческих процессах организации. При этом существенным фактором успеха является выбор правильного инструментария: развитие методологии анализа больших данных позволило сформировать широкий спектр аналитических методов обработки больших данных, включающий описательную, диагностическую, предсказательную и предписательную аналитику.

В настоящее время компании-лидеры используют специализированные инструменты на всех стадиях жизненного цикла работы с большими данными (генерация, сбор, хранение, предобработка/обработка, анализ и визуализация), а аналитику больших данных применяют практически для всех видов деловых процессов.

Этим продиктован выбор следующих индикаторов для размерности «Инструменты и аналитика» модели зрелости BD4DE-MM.

1. Использование специализированных инструментов работы с большими данными.
2. Регулярность использования аналитики больших данных в организации.
3. Направления использования аналитики больших данных в организации.

2.5 Инфраструктура и безопасность

Наличие развитой цифровой инфраструктуры, удовлетворяющей нарастающие потребности в генерации, обработке, передаче, хранении и анализе больших данных, является одним из важнейших факторов, определяющих эффективность работы с большими данными в организации. В предлагаемой модели BD4DE-ММ уровень зрелости инфраструктуры оценивается с двух точек зрения: (1) наличия необходимой инфраструктуры для работы с большими данными в организации, (2) возможности масштабируемости этой инфраструктуры.

Информационная безопасность является существенным фактором, который влияет на использование больших данных в организациях, поскольку, с одной стороны, может выступать препятствием (в силу своей сложности, обязательности соблюдения законодательства о защите персональных данных, недостаточной развитости и чрезмерной затратности), а с другой — влиять на повышение доверия сотрудников и контрагентов организации к работе с большими данными (при условии обеспечения высокого уровня безопасности).

Активное развитие и использование технологий хранения и анализа больших данных формирует множество вызовов при обеспечении информационной безопасности. Сверхбольшие объемы данных, их разнообразие и высокая скорость генерации и обработки данных порождают дополнительные риски, зачастую не возникающие в процессе использования традиционных ИКТ, и требуют специализированного инструментария. С учетом международного опыта мониторинга и оценки информационной безопасности, закрепленных, в частности, в международном стандарте ISO/IEC 20547-4:2020 «Информационные технологии. Эталонная архитектура больших данных. Часть 4. Безопасность и защита персональных данных» [44], в предлагаемой модели зрелости BD4DE-ММ оцениваются ключевые составляющие обеспечения информационной безопасности при работе с большими данными в организациях: регламентация процессов обеспечения информационной безопасности, наличие необходимых навыков у сотрудников, обеспеченность организации специалистами по информационной безопасности работы с большими данными, использование специализированных инструментов для обеспечения безопасности на различных стадиях жизненного цикла работы с большими данными.

На основе описанного выше подхода были выбраны следующие индикаторы для характеристики размерности «Инфраструктура и безопасность» модели зрелости BD4DE-ММ.

1. Наличие инфраструктуры для работы с большими данными в организации.
2. Масштабируемость инфраструктуры для работы с большими данными в организации.
3. Регламентирование процессов обеспечения доверия и безопасности работы с большими данными в организации.
4. Наличие у сотрудников организации навыков по обеспечению доверия и безопасности работы с большими данными.
5. Потребности организации в специалистах по обеспечению доверия и безопасности работы с большими данными.
6. Использование в организации специализированных инструментов обеспечения доверия и безопасности при работе с большими данными.

2.6 Организация работы с большими данными

В последние годы организационные аспекты деятельности все чаще анализируются в терминах организационного или структурного капитала. Это связано с тем, что оптимальная организация деловых процессов и развитая корпоративная культура существенно влияют на функционирование организации и могут рассматриваться как важнейший нематериальный актив, вложение в который окупается ростом эффективности использования остальных ресурсов. Важной составляющей организационного капитала является собственно система управления.

При определении уровня зрелости работы с большими данными в организации необходимо, прежде всего, оценить такие аспекты управления работой с большими данными, как наличие руководителя, отвечающего за работу с (большими) данными, наличие системы мониторинга работы с большими данными для обеспечения обратной связи в системе управления, а также использование современных методов проектного управления при работе с большими данными. Кроме того, оценивается наличие специализированного подразделения по работе с большими данными и центра компетенций в этой сфере. К институциональным факторам также относится регламентация процессов работы с большими данными в организации.

Таким образом размерность «Организация работы» в модели зрелости BD4DE-MM характеризуется следующими индикаторами.

1. Регламентирование процессов работы с большими данными в организации.
2. Наличие системы мониторинга работы с большими данными в организации.
3. Наличие в организации руководителя, отвечающего за работу с большими данными.
4. Наличие в организации структурного подразделения, отвечающего за работу с большими данными.
5. Наличие в организации центра компетенций по работе с большими данными.
6. Управление проектами по работе с большими данными в организации.

2.7 Воздействие

Главной целью применения технологий работы с большими данными в организациях является получение дивидендов: экономических и неэкономических (в том числе - социальных) эффектов. Отсутствие ощутимого воздействия на деятельность организации ставит под вопрос целесообразность вложений в работу с большими данными и свидетельствует о низком уровне зрелости такой работы. Поэтому оценка реально достигнутых эффектов работы с большими данными в организации является важной характеристикой зрелости.

Как отмечено выше, выделяют две группы эффектов использования цифровых технологий – экономические и социальные (точнее сказать – неэкономические), хотя такое разбиение и носит достаточно условный характер. В отдельных случаях или для некоторых видов деловых процессов это можно сделать целенаправленно, однако иногда можно оценить лишь обобщенные эффекты, воздействующие на организацию в целом. Поэтому в модель зрелости BD4DE-MM для характеристики размерности «Воздействие» были включены следующие интегральные индикаторы.

1. Воздействие работы с большими данными на деятельность организации в целом.
2. Экономические эффекты от использования больших данных в организации.
3. Неэкономические эффекты от использования больших данных в организации.

Заключение

В данной статье представлена модель зрелости BD4DE-MM работы с большими данными в организациях, учитывающая основные тенденции в области построения моделей зрелости в общем контексте цифровой экономики. Разработанная модель зрелости может служить как при оценке зрелости организации внешними экспертами, так и в режиме самооценки самими сотрудниками. Использование модели BD4DE-MM также позволяет проводить оценку зрелости работы с большими данными как на уровне отдельной сферы деятельности отрасли, так и для страны в целом. Соответствующие результаты могут быть использованы и руководителями отдельных организаций, и отраслевыми органами управления для принятия решений по выработке соответствующих решений по развитию.

Дальнейшие исследования могут быть продолжены в направлении дополнения перечня размерностей и уточнения отдельных аспектов, расширения состава используемых при работе с большими данными технологий (например, технологий машинного или глубокого обучения, более продвинутых технологий искусственного интеллекта). Аналогичный подход может быть также применен для построения моделей зрелости использования других цифровых технологий в организациях или отраслях.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Литература

1. Katz, Raul. Social and economic impact of digital transformation on the economy. International Telecommunications Union, July 2017. 41 p.
2. World Bank. 2016. World Development Report 2016: Digital Dividends. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-0671-1>
3. World Bank Group. 2018. Competing in the Digital Age: Policy Implications for the Russian Federation. Russia Digital Economy Report; World Bank, Washington, DC. World Bank. URL: <https://openknowledge.worldbank.org/handle/10986/30584> (дата обращения 01.06.2021)
4. Оценка уровня развития цифровой экономики: Методика ДЕСА оценки уровня развития цифровой экономики. URL: <http://deca.iis.ru> (дата обращения 15.05.2021)
5. Capability Maturity Model Integration (CMMI). URL: <https://cmmiinstitute.com/cmmi/> (дата обращения 01.06.2021)
6. CMMI Institute, URL: <https://cmmiinstitute.com/> (дата обращения 01.06.2021)
7. ISO/IEC 33001:2015 Information technology – Process assessment – Concepts and terminology
8. ISO/IEC 33004:2015 Information technology – Process assessment – Requirements for process reference, process assessment and maturity models.
9. ГОСТ Р ИСО/МЭК 33001–2017 Информационные технологии – Оценка процесса – Понятия и терминология
10. ГОСТ Р ИСО/МЭК 33004–2017 Информационные технологии. Оценка процесса. Требования к эталонным моделям процесса, моделям оценки процесса и моделям зрелости
11. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
12. Diogo Proença, José Borbinha. Maturity Models for Data and Information Management. DOI: https://doi.org/10.1007/978-3-030-00066-0_7 (дата обращения 01.06.2021)
13. S. Eybers, M.J. Hattingh. Critical Success Factor Categories for Big Data: A Preliminary Analysis of the Current Academic Landscape. URL: https://www.researchgate.net/publication/317344777_Critical_Success_Factor_Categories_for_Big_Data_A_Preliminary_Analysis_of_the_Current_Academic_Landscape (дата обращения 19.11.2020)
14. Jorne Mathijs Evers. Critical Success Factors of Business Intelligence and Big Data Analysis Utilizing Hidden Business Value from Big Data for Business Intelligence. Master Thesis. URL: https://www.researchgate.net/publication/317598149_Critical_factors_for_Business_Intelligence_Success (дата обращения 19.11.2020)
15. Patrick Cato, P. Golzer, Walter Demmelhuber. An investigation into the implementation factors affecting the success of big data systems. URL: <https://www.semanticscholar.org/paper/An-investigation-into-the-implementation-factors-of-Cato-Golzer/88ecb1dc28b7191be822f95947933cff2a9a00b7> (дата обращения 19.11.2020)
16. Hong-Mei Chen, Roland Schütz, Rick Kazman, Florian Matthes. Amazon in the Air: Innovating with Big Data at Lufthansa. URL: https://www.researchgate.net/publication/301281721_Amazon_in_the_Air_Innovating_with_Big_Data_at_Lufthansa (дата обращения 19.11.2020)
17. Koronios, Andy, Gao, Jing, Selle, Sven. Big Data Project Success – A Meta Analysis. URL: <https://www.semanticscholar.org/paper/Big-Data-Project-Success-a-Meta-Analysis-Koronios-Gao/615937f9ce8ab8b050a2da5a8c9fb0e5566591b1> (дата обращения 19.11.2020)
18. Jeffrey S. Saltz, Ivan Shamshurin. Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project’s Success. URL: <https://www.semanticscholar.org/paper/Big-data-team-process-methodologies%3A-A-literature-a-Saltz-Shamshurin/b25cc739b7f0fb9db1154a274f52c5e3ace0de5f> (дата обращения 19.11.2020)
19. Cecilia Adrian, Serdang Selangor, Rusli Abdullah, Rodziah Atan, Yusmadi Yah Jusoh. Factors Influencing to the Implementation Success of Big Data Analytics: A Systematic Literature Review. URL: <https://www.semanticscholar.org/paper/Factors-influencing-to-the-implementation-success-A-Adrian-Abdullah/feabd664c013bc8db3a151e512685f667f2a3d4c> (дата обращения 19.11.2020)
20. Jing Gao, Andy Koronios. Towards A Process View on Critical Success Factors in Big Data Analytics Projects. URL: <https://www.semanticscholar.org/paper/Towards-A-Process-View-on-Critical->

- [Success-Factors-Gao-Koronios/247bfe6fa3365d74bd98c2c460785d62c3d7561d](https://www.researchgate.net/publication/322185870) (дата обращения 19.11.2020)
21. Bruno Muniz Félix, Elaine Tavares, Ney Wagner Freitas Cavalcante. Critical success factors for Big Data adoption in the virtual retail: Magazine Luiza case study. URL: <https://www.researchgate.net/publication/322185870> [Critical success factors for Big Data adoption in the virtual retail Magazine Luiza case study](https://www.researchgate.net/publication/322185870) (дата обращения 19.11.2020)
 22. Mohanad Halaweh, Ahmed E. Massry. Conceptual Model for Successful Implementation of Big Data in Organizations. URL: <https://www.semanticscholar.org/paper/Conceptual-Model-for-Successful-Implementation-of-Halaweh-Massry/ac87fa6782c86a29d49819373328fc53e06d2048> (дата обращения 19.11.2020)
 23. Liliana Fernández Gómez & Richard Heeks. Measuring the Barriers to Big Data for Development: Design-Reality Gap Analysis in Colombia's Public Sector. URL: <https://www.researchgate.net/publication/335162375> [Measuring the Barriers to Big Data for Development Design-Reality Gap Analysis in Colombia's Public Sector](https://www.researchgate.net/publication/335162375) (дата обращения 19.11.2020)
 24. Moon-Koo Kim, Jong-Hyun Park. Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. URL: <http://metatoc.com/papers/77110-identifying-and-prioritizing-critical-factors-for-promoting-the-implementation-and-usage-of-big-data-in-healthcare> (дата обращения 19.11.2020)
 25. Anke Schüll and Natalia Maslan. On the Adoption of Big Data Analytics: Interdependencies of Contextual Factors. URL: <https://www.researchgate.net/publication/324044546> [On the Adoption of Big Data Analytics Interdependencies of Contextual Factors](https://www.researchgate.net/publication/324044546) (дата обращения 19.11.2020)
 26. Yichuan Wang, LeeAnn Kung, Terry Anthony Byrd. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations URL: <https://www.ehdc.org/sites/default/files/resources/files/big%20data%20analytics.pdf> (дата обращения 19.11.2020)
 27. Andreas Schumacher, Selim Erol, Wilfried Sihm. A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. URL: <https://www.researchgate.net/publication/307620022> [A Maturity Model for Assessing Industry 4.0 Readiness and aturity of Manufacturing Enterprises](https://www.researchgate.net/publication/307620022) (дата обращения 19.11.2020)
 28. Andreas Schumacher, Tanja Nemetha, Wilfried Sihm. Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises. URL: <https://www.semanticscholar.org/paper/Roadmapping-towards-industrial-digitalization-based-Schumacher-Nemeth/778742f70e1d2f514202efb9b225e2511491d12> (дата обращения 19.11.2020)
 29. Шу, Г., Андерл, Р., Гауземайер, Ю., тен Хомпель, М., Вальстер, В. (и др.): Индекс зрелости Индустрии 4.0 – Управление цифровым преобразованием компаний (acatech ИССЛЕДОВАНИЕ), Munich: Herbert Utz Verlag 2017. URL: https://www.acatech.de/wp-content/uploads/2018/03/acatech_STUDIE_rus_Maturity_Index_WEB.pdf (дата обращения 19.11.2020)
 30. Industry 4.0 Readiness Online Self-Check for Businesses. URL: <https://www.industrie40-readiness.de/> (дата обращения 19.11.2020)
 31. TDWI IoT Data Readiness Assessment. URL: <https://tdwi.org/pages/assessments/arch-all-iot-data-readiness-assessment.aspx> (дата обращения 19.11.2020)
 32. Data Orchard. Data Maturity Framework. URL: <https://www.dataorchard.org.uk/resources/data-maturity-framework> (дата обращения 19.11.2020)
 33. Gartner's Enterprise Information Management Maturity Model. URL: <https://www.gartner.com/en/documents/3236418> (дата обращения 01.06.2021)
 34. TDWI. A Guide to Achieving Big Data Analytics Maturity (v 1.1.2018). URL: <https://tdwi.org/whitepapers/2018/01/adv-all-ms-a-guide-to-achieving-big-data-analytics-maturity.aspx> (дата обращения 19.11.2020)
 35. Hortonworks Big Data Maturity Model. A Hortonworks White Paper. URL: <http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf> (дата обращения 19.11.2020)
 36. IDC MaturityScape Benchmark: Big Data and Analytics in the United States. URL: <https://tdwi.org/whitepapers/2016/06/cloudera-idc-big-data-and-analytics-in-the-us.aspx> (дата обращения 19.11.2020)

37. IBM Big Data & Analytics Maturity Model. URL: https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en (дата обращения 19.11.2020)
38. TM Forum. AI Maturity Model Toolkit. URL: <https://www.tmforum.org/resources/toolkit/ai-maturity-model-toolkit/> (дата обращения 01.10.2021)
39. TDWI Analytics Maturity Model. URL: <https://tdwi.org/analytics-assessment> (дата обращения 19.11.2020)
40. TDWI IoT Readiness Guide: Interpreting Your Assessment Score. URL: <https://tdwi.org/whitepapers/2016/08/tdwi-iot-readiness-guide-ibm.aspx> (дата обращения 19.11.2020)
41. TDWI Advanced Analytics Maturity Model Assessment. URL: <https://tdwi.org/whitepapers/2018/04/adv-all-tdwi-advanced-analytics-maturity-model-guide.aspx> (дата обращения 19.11.2020).
42. TDWI Hadoop Readiness Guide: Interpreting Your Assessment Score. URL: <https://tdwi.org/whitepapers/2015/10/tdwi-hadoop-readiness-guide.aspx> (дата обращения 19.11.2020)
43. TDWI Self-Service Analytics Maturity Model Guide: URL: <https://tdwi.org/whitepapers/2017/10/bi-all-tdwi-self-service-analytics-maturity-model-guide.aspx> (дата обращения 19.11.2020)
44. ISO/IEC 20547-4:2020 Information technology – Big data reference architecture – Part 4: Security and privacy

BIG DATA FOR DIGITAL ECONOMY MATURITY MODEL

Ershov, Peter Sergeevich

*Institute of the Information Society, head of Directorate of regional programs
Moscow, Russian Federation
peter.ershov@iis.ru*

Katin, Alexander Vladimirovich

*Institute of the Information Society, head of Directorate of sectoral programs
Plekhanov Russian University of Economics, IIS-based Digital economy department, senior lecturer
Moscow, Russian Federation
alexander.katin@iis.ru*

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Shaposhnik, Sergei Borisovich

*Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher
Petrozavodsk, Russian Federation
sergei.shaposhnik@gmail.com*

Abstract

A maturity model BD4DE-MM for assessment of big data usage in an organization is proposed. The analysis of the existing models of digital maturity for organization is carried out, the following main dimensions are highlighted: strategy and regulation; human resources and leadership; data; tools and analytics; infrastructure and cybersecurity; governance; impact. Indicators are proposed that allow for each dimension to assess on a 5-point scale the maturity level of big data usage in an organization.

Keywords

maturity model; big data; key success factors; use of big data; big data analytics; BD4DE-MM; Big Data for Digital Economy Maturity Model; strategy; regulation; human resources; leadership; data; tools; infrastructure; security; governance; impact

References

1. Katz, Raul. Social and economic impact of digital transformation on the economy. International Telecommunications Union, July 2017. 41 p.
2. World Bank. 2016. World Development Report 2016: Digital Dividends. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-0671-1>
3. World Bank Group. 2018. Competing in the Digital Age: Policy Implications for the Russian Federation. Russia Digital Economy Report; World Bank, Washington, DC. World Bank. URL: <https://openknowledge.worldbank.org/handle/10986/30584> (accessed on 01.06.2021)
4. Otsenka urovnya razvitiya tsifrovoy ekonomiki: Metodika DECA otsenki urovnya razvitiya tsifrovoy ekonomiki. URL: <http://deca.iis.ru> (accessed on 15.05.2021)
5. Capability Maturity Model Integration (CMMI). URL: <https://cmmiinstitute.com/cmmi/> (accessed on 01.06.2021)
6. CMMI Institute, URL: <https://cmmiinstitute.com/> (accessed on 01.06.2021)
7. ISO/IEC 33001:2015 Information technology – Process assessment – Concepts and terminology
8. ISO/IEC 33004:2015 Information technology – Process assessment – Requirements for process reference, process assessment and maturity models.
9. GOST R ISO/MEK 33001–2017 Informatsionnyye tekhnologii – Otsenka protsesssa – Ponyatiya i terminologiya

10. 10. GOST R ISO/MEK 33004–2017 Informatsionnyye tekhnologii. Otsenka protsessa. Trebovaniya k etalonnym modelyam protsessa, modelyam otsenki protsessa i modelyam zrelosti
11. World Bank. 2021. World Development Report 2021: Data for Better Lives. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
12. Diogo Proença, José Borbinha. Maturity Models for Data and Information Management. DOI: https://doi.org/10.1007/978-3-030-00066-0_7 (accessed on 01.06.2021)
13. S. Eybers, M.J. Hattingh. Critical Success Factor Categories for Big Data: A Preliminary Analysis of the Current Academic Landscape. URL: https://www.researchgate.net/publication/317344777_Critical_Success_Factor_Categories_for_Big_Data_A_Preliminary_Analysis_of_the_Current_Academic_Landscape (accessed on 19.11.2020)
14. Jorne Mathijs Evers. Critical Success Factors of Business Intelligence and Big Data Analysis Utilizing Hidden Business Value from Big Data for Business Intelligence. Master Thesis. URL: https://www.researchgate.net/publication/317598149_Critical_factors_for_Business_Intelligence_Success (accessed on 19.11.2020)
15. Patrick Cato, P. Golzer, Walter Demmelhuber. An investigation into the implementation factors affecting the success of big data systems. URL: <https://www.semanticscholar.org/paper/An-investigation-into-the-implementation-factors-of-Cato-Golzer/88ecb1dc28b7191be822f95947933cff2a9a00b7> (accessed on 19.11.2020)
16. Hong-Mei Chen, Roland Schütz, Rick Kazman, Florian Matthes. Amazon in the Air: Innovating with Big Data at Lufthansa. URL: https://www.researchgate.net/publication/301281721_Amazon_in_the_Air_Innovating_with_Big_Data_at_Lufthansa (accessed on 19.11.2020)
17. Koronios, Andy, Gao, Jing, Selle, Sven. Big Data Project Success – A Meta Analysis. URL: <https://www.semanticscholar.org/paper/Big-Data-Project-Success-a-Meta-Analysis-Koronios-Gao/615937f9ce8ab8b050a2da5a8c9fb0e5566591b1> (accessed on 19.11.2020)
18. Jeffrey S. Saltz, Ivan Shamshurin. Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project’s Success. URL: <https://www.semanticscholar.org/paper/Big-data-team-process-methodologies%3A-A-literature-a-Saltz-Shamshurin/b25cc739b7f0fb9db1154a274f52c5e3ace0de5f> (accessed on 19.11.2020)
19. Cecilia Adrian, Serdang Selangor, Rusli Abdullah, Rodziah Atan, Yusmadi Yah Jusoh. Factors Influencing to the Implementation Success of Big Data Analytics: A Systematic Literature Review. URL: <https://www.semanticscholar.org/paper/Factors-influencing-to-the-implementation-success-A-Adrian-Abdullah/feabd664c013bc8db3a151e512685f667f2a3d4c> (accessed on 19.11.2020)
20. Jing Gao, Andy Koronios. Towards A Process View on Critical Success Factors in Big Data Analytics Projects. URL: <https://www.semanticscholar.org/paper/Towards-A-Process-View-on-Critical-Success-Factors-Gao-Koronios/247bfe6fa3365d74bd98c2c460785d62c3d7561d> (accessed on 19.11.2020)
21. Bruno Muniz Félix, Elaine Tavares, Ney Wagner Freitas Cavalcante. Critical success factors for Big Data adoption in the virtual retail: Magazine Luiza case study. URL: https://www.researchgate.net/publication/322185870_Critical_success_factors_for_Big_Data_adoption_in_the_virtual_retail_Magazine_Luiza_case_study (accessed on 19.11.2020)
22. Mohanad Halaweh, Ahmed E. Massry. Conceptual Model for Successful Implementation of Big Data in Organizations. URL: <https://www.semanticscholar.org/paper/Conceptual-Model-for-Successful-Implementation-of-Halaweh-Massry/ac87fa6782c86a29d49819373328fc53e06d2048> (accessed on 19.11.2020)
23. Liliana Fernández Gómez & Richard Heeks. Measuring the Barriers to Big Data for Development: Design-Reality Gap Analysis in Colombia’s Public Sector. URL: https://www.researchgate.net/publication/335162375_Measuring_the_Barriers_to_Big_Data_for_Development_Design-Reality_Gap_Analysis_in_Colombia's_Public_Sector (accessed on 19.11.2020)
24. Moon-Koo Kim, Jong-Hyun Park. Identifying and prioritizing critical factors for promoting the implementation and usage of big data in healthcare. URL: <http://metatoc.com/papers/77110-identifying-and-prioritizing-critical-factors-for-promoting-the-implementation-and-usage-of-big-data-in-healthcare> (accessed on 19.11.2020)
25. Anke Schüll and Natalia Maslan. On the Adoption of Big Data Analytics: Interdependencies of Contextual Factors. URL: https://www.researchgate.net/publication/324044546_On_the_Adoption_of_Big_Data_Analytics_Interdependencies_of_Contextual_Factors (accessed on 19.11.2020)

26. Yichuan Wang, LeeAnn Kung, Terry Anthony Byrd. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations URL: <https://www.ehdc.org/sites/default/files/resources/files/big%20data%20analytics.pdf> (accessed on 19.11.2020)
27. Andreas Schumacher, Selim Erol, Wilfried Sihm. A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. URL: https://www.researchgate.net/publication/307620022_A_Maturity_Model_for_Assessing_Industry_40_Readiness_and_maturity_of_Manufacturing_Enterprises (accessed on 19.11.2020)
28. Andreas Schumacher, Tanja Nemeth, Wilfried Sihm. Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises. URL: <https://www.semanticscholar.org/paper/Roadmapping-towards-industrial-digitalization-based-Schumacher-Nemeth/778742f70e1d2f5142020efb9b225e2511491d12> (accessed on 19.11.2020)
29. Shu, G., Anderl, R., Gauzemayyer, YU., ten Khompel', M., Val'ster, V. (i dr.): Indeks zrelosti Industrii 4.0 – Upravleniye tsifrovym preobrazovaniyem kompaniy (acatech ISSLEDOVANIYE), Munich: Herbert Utz Verlag 2017. URL: https://www.acatech.de/wp-content/uploads/2018/03/acatech_STUDIE_rus_Maturity_Index_WEB.pdf (accessed on 19.11.2020)
30. Industry 4.0 Readiness Online Self-Check for Businesses. URL: <https://www.industrie40-readiness.de/> (accessed on 19.11.2020)
31. TDWI IoT Data Readiness Assessment. URL: <https://tdwi.org/pages/assessments/arch-all-iot-data-readiness-assessment.aspx> (accessed on 19.11.2020)
32. Data Orchard. Data Maturity Framework. URL: <https://www.dataorchard.org.uk/resources/data-maturity-framework> (accessed on 19.11.2020)
33. Gartner's Enterprise Information Management Maturity Model. URL: <https://www.gartner.com/en/documents/3236418> (accessed on 01.06.2021)
34. TDWI. A Guide to Achieving Big Data Analytics Maturity (v 1.1.2018). URL: <https://tdwi.org/whitepapers/2018/01/adv-all-ms-a-guide-to-achieving-big-data-analytics-maturity.aspx> (accessed on 19.11.2020)
35. Hortonworks Big Data Maturity Model. A Hortonworks White Paper. URL: <http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf> (accessed on 19.11.2020)
36. IDC MaturityScape Benchmark: Big Data and Analytics in the United States. URL: <https://tdwi.org/whitepapers/2016/06/cloudera-idc-big-data-and-analytics-in-the-us.aspx> (accessed on 19.11.2020)
37. IBM Big Data & Analytics Maturity Model. URL: https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en (accessed on 19.11.2020)
38. TM Forum. AI Maturity Model Toolkit. URL: <https://www.tmforum.org/resources/toolkit/ai-maturity-model-toolkit/> (accessed on 01.10.2021)
39. TDWI Analytics Maturity Model. URL: <https://tdwi.org/analytics-assessment> (accessed on 19.11.2020)
40. TDWI IoT Readiness Guide: Interpreting Your Assessment Score. URL: <https://tdwi.org/whitepapers/2016/08/tdwi-iot-readiness-guide-ibm.aspx> (accessed on 19.11.2020)
41. TDWI Advanced Analytics Maturity Model Assessment. URL: <https://tdwi.org/whitepapers/2018/04/adv-all-tdwi-advanced-analytics-maturity-model-guide.aspx> (accessed on 19.11.2020).
42. TDWI Hadoop Readiness Guide: Interpreting Your Assessment Score. URL: <https://tdwi.org/whitepapers/2015/10/tdwi-hadoop-readiness-guide.aspx> (accessed on 19.11.2020)
43. TDWI Self-Service Analytics Maturity Model Guide: URL: <https://tdwi.org/whitepapers/2017/10/bi-all-tdwi-self-service-analytics-maturity-model-guide.aspx> (accessed on 19.11.2020)
44. ISO/IEC 20547-4:2020 Information technology – Big data reference architecture – Part 4: Security and privacy

Технологии информационного общества

ИНДУСТРИЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Малахов Вадим Александрович

Кандидат исторических наук

Институт истории естествознания и техники им. С. И. Вавилова РАН, Центр истории организации науки и науковедения, старший научный сотрудник

Москва, Российская Федерация

yasonbh@mail.ru

Хохлов Юрий Евгеньевич

Кандидат физико-математических наук, доцент

Институт развития информационного общества, председатель совета директоров

РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО

Москва, Российская Федерация

yuri.hohlov@iis.ru

Шапошник Сергей Борисович

Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития, старший научный сотрудник

Петрозаводск, Российская Федерация

sergei.shaposhnik@gmail.com

Аннотация

В статье предложен подход, концептуальная схема и показатели для мониторинга индустрии работы с большими данными, под которой понимается совокупность компаний, производящих технологии, решения и услуги в области работы с большими данными. Концептуальная схема мониторинга основана на ролевой модели работы с большими данными, представленной в стандарте эталонной архитектуры больших данных Международной организации по стандартизации (ISO). В систему мониторинга включены 9 показателей, характеризующих масштабы, конкурентоспособность и потенциал индустрии работы с большими данными. Измерение показателя, связанного с числом стартапов в области индустрии работы с большими данными и основанного на базе данных Crunchbase, показало существенное отставание России по этому показателю от развитых стран (0,52% от всех стартапов в этой области).

Ключевые слова

большие данные; технологии работы с большими данными; индустрия работы с большими данными; мониторинг

Введение

Быстрый рост объемов данных, доступных для использования, а также развитие методов извлечения полезной информации из этих массивов данных приводят к росту спроса на технологии работы с большими данными в различных сферах деятельности. В ответ на этот спрос формируется новый сектор экономики – индустрия работы с большими данными, который объединяет поставщиков технологий, решений и услуг работы с большими данными. Развитие в России собственной индустрии работы с большими данными важно с нескольких точек зрения, включая решение задач создания новых высокопроизводительных рабочих мест, импортозамещения и обеспечения информационной безопасности при использовании больших данных в чувствительных сферах (оборона, национальная безопасность и др.).

© Малахов В.А., Хохлов Ю.Е., Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_278

В данной статье предлагается комплексная методология мониторинга и оценки уровня отечественного производства технологий работы с большими данными на основе системы показателей, позволяющей измерять текущие масштабы индустрии работы с большими данными, ее глобальную конкурентоспособность и потенциал развития.

1 Определение предметной области мониторинга и оценки

Предметом мониторинга является состояние и потенциал развития сектора отечественной экономики, производящего технологии, решения и услуги, связанные с большими данными (далее – индустрии работы с большими данными).

С точки зрения классификации видов экономической деятельности (в России это ОКВЭД2) индустрия работы с большими данными является частью цифрового сектора экономики, под которым мы будем понимать группировку видов экономической деятельности, входящих в ИКТ-сектор и сектор контента и СМИ [1,2].

Наиболее удобным подходом к мониторингу индустрии работы с большими данными было бы формирование группировки кодов экономической деятельности, связанных с работой с большими данными, что позволило бы получать для этого из официальной статистики значения показателей, характеризующих масштабы, инновационную активность и т.д. Подобным образом, например, осуществляется мониторинг развития цифрового сектора в целом: доступные показатели, собираемые в рамках федерального статистического наблюдения, берутся в разрезе кодов ОКВЭД2, входящих в цифровой сектор экономики, и агрегируются для характеристики сектора. К сожалению, использование аналогичного подхода для такого специфического сектора экономики как производство технологий, решений и услуг, связанных с большими данными невозможно – коды классификатора видов экономической деятельности слишком «грубы» и широки, чтобы отразить нужные виды деятельности. В качестве примера можно привести тот факт, что для разработки программного обеспечения в ОКВЭД2 есть только один код (62.01 Разработка компьютерного программного обеспечения), который охватывает все виды программного обеспечения и не позволяет, например, выделить из них разработку технологий или приложений для работы с большими данными.

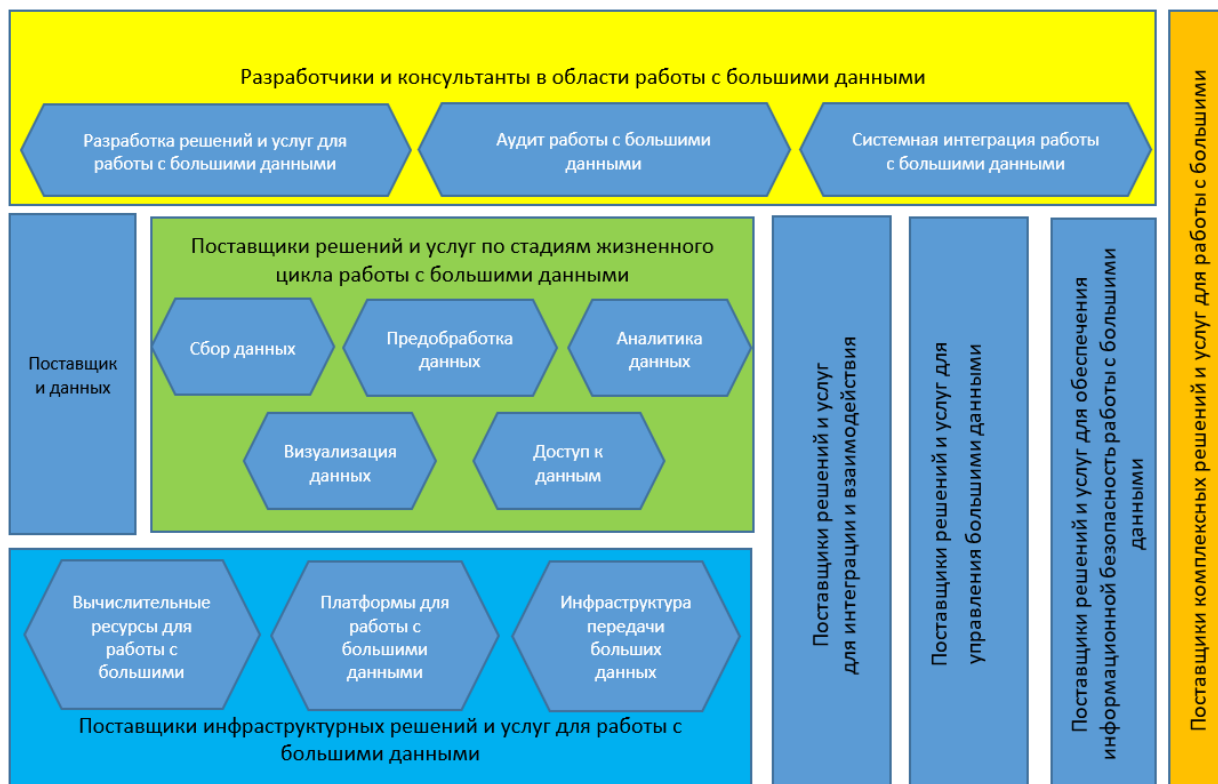


Рисунок 1. Производители (поставщики) технологий, решений и услуг для работы с большими данными

В этой ситуации для выделения индустрии работы с большими данными как сектора экономики был разработан подход, основанный на стандартах Международной организации по стандартизации (ISO), связанных с технологиями работы с большими данными, в частности, учитывающий стадии жизненного цикла работы с большими данными, зафиксированного в стандартах эталонной архитектуры больших данных [3]. В основу классификатора видов деятельности, которые предлагается включить в индустрию работы с большими данными, положена ролевая модель оперирования большими данными по жизненному циклу, а также обеспечивающие его вспомогательные и интеграционные виды деятельности. В рамках этого подхода к компаниям «индустрии работы с большими данными» предлагается относить поставщиков технологий, решений и услуг, схематично представленных на рисунке 1.

Указанные виды деятельности попадают в более широкие по охвату коды Общероссийского классификатора видов экономической деятельности (ОКВЭД2), представленные ниже в таблице 1.

Таблица 1. Соотношение классификации компаний индустрии работы с большими данными и кодов ОКВЭД2

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<i>Разработчики и консультанты в области работы с большими данными</i>	
<p>Разработка решений и технологий оказания услуг для работы с большими данными Включает формирование требований, планирование (концептуализацию), проектирование, разработку, тестирование технологий работы с большими данными и, при необходимости, передачу другому лицу технологической документации и необходимых знаний Включает формирование требований, планирование создания, проектирование, разработку, тестирование и подготовку к вводу в эксплуатацию систем для работы с большими данными и/или сервисов на их основе</p>	<p>62.01 Разработка компьютерного программного обеспечения 62.02.1 Деятельность по планированию, проектированию компьютерных систем 62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>
<p>Аудит массивов больших данных, решений и услуг для работы с большими данными Включает проведение аудита процессов и систем для работы с большими данными, связанного с проверкой достоверности источников и качества данных, операций с массивами больших данных, выполнением требований безопасности и защиты персональных данных, а также проверке соответствия заданному набору критериев аудита</p>	<p>62.02.2 Деятельность по обследованию и экспертизе компьютерных систем 63.1 Деятельность по обработке данных, предоставление услуг по размещению информации, деятельность порталов в информационно-коммуникационной сети Интернет</p>
<p>Системная интеграция работы с большими данными Включает формирование требований к решениям и услугам для работы с большими данными, включая требования к политике, системе (стратегического) управления, архитектуре, ресурсам, безопасности, качеству данных и требования бизнеса, деятельность по практической реализации разработанных требований, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>62.01 Разработка компьютерного программного обеспечения 62.02.1 Деятельность по планированию, проектированию компьютерных систем 62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<i>Поставщики больших данных</i>	
<p>Поставка данных</p> <p>Включает передачу, в том числе путем предоставления доступа, другому лицу массивов больших данных и прав на их использование, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>63.1 Деятельность по обработке данных, предоставление услуг по размещению информации, деятельность порталов в информационно-коммуникационной сети Интернет</p> <p>63.9 Деятельность в области информационных услуг прочая</p> <p>63.99 Деятельность информационных служб прочая, не включенная в другие группировки</p>
<i>Поставщики решений и услуг по стадиям жизненного цикла работы с большими данными</i>	
<p>Сбор (комплектование) данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для сбора данных и/или сервисов на их основе, связанных с выбором источников данных, приемом массивов больших данных, их регистрацией и буферизацией для последующих действий, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>58.29 Издание прочих программных продуктов</p> <p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.02.1 Деятельность по планированию, проектированию компьютерных систем</p> <p>62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>
<p>Предобработка данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для предварительной обработки данных и/или сервисов на их основе, связанных с подготовкой полученных массивов больших данных к действиям над ними (управлению, преобразованию, обработке, анализу, визуализации, хранению, удалению и иным действиям), а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<p>Аналитика данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для аналитики данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<p>Визуализация данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для визуализации данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<p>Доступ к данным</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для управления доступом к массивам больших данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<i>Поставщики инфраструктурных решений и услуг для работы с большими данными</i>	
<p>Предоставление платформенных решений для работы с большими данными</p> <p>Включает предоставление в пользование (аренду), или передачу другому лицу прав на использование платформенных решений для работы с большими данными и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p> <p>26.30 Производство коммуникационного оборудования</p>
<p>Предоставление вычислительных ресурсов для работы с большими данными</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для работы с большими данными и/или сервисов на их основе, связанных с вычислительными средствами (например, компьютерами, хранилищами данных) и/или физической средой (компьютерными залами, стойко-местами, электропитанием, кондиционерами и т.д.), обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.80 Производство незаписанных магнитных и оптических технических носителей информации</p> <p>61.10.3 Деятельность по предоставлению услуг по передаче данных и услуг доступа к информационно-коммуникационной сети Интернет</p>
<p>Предоставление телекоммуникационных решений для работы с большими данным</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для работы с большими данными и/или сервисов на их основе, связанных с телекоммуникационными средствами для передачи массивов больших данных, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.03 Деятельность по управлению компьютерным оборудованием</p> <p>63.11 Деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность</p> <p>63.12 Деятельность веб-порталов</p>
<i>Поставщики решений и услуг, обеспечивающих работу с большими данными</i>	
<p>Поставка решений и услуг для интеграции и взаимодействия систем для работы с большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для интеграции и взаимодействия систем оперирования большими данными и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p>

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<p>Поставка решений и услуг для управления большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для управления массивами больших данных и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.30 Производство коммуникационного оборудования</p> <p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.03 Деятельность по управлению компьютерным оборудованием</p>
<p>Поставка решений и услуг для обеспечения информационной безопасности работы с большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для обеспечения информационной безопасности систем оперирования массивами больших данных и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<p><i>Поставщики комплексных решений и услуг для работы с большими данными</i></p>	
<p>Поставка комплексных решений и услуг для работы с большими данными</p> <p>Включает несколько перечисленных выше одновременно выполняемых видов деятельности по работе с большими данными</p>	<p>61.10.3 Деятельность по предоставлению услуг по передаче данных и услуг доступа к информационно-коммуникационной сети Интернет</p> <p>62 Разработка компьютерного программного обеспечения, консультационные услуги в данной области и другие сопутствующие услуги</p> <p>63 Деятельность в области информационных технологий</p> <p>58.29 Издание прочих программных продуктов</p> <p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p> <p>26.30 Производство коммуникационного оборудования</p> <p>26.80 Производство незаписанных магнитных и оптических технических носителей информации</p>

Таким образом, операционализируя данное выше определение индустрии работы с большими данными как сектора экономики, можно сказать, что это совокупность компаний, деятельность которых описывается кодами ОКВЭД2, входящими в цифровой сектор экономики (таблица 1) и выделяемыми в соответствии с видами деятельности, указанными на рисунке 1.

Аналогичный подход был реализован в проекте Европейской комиссии «Европейский рынок данных», в котором решалась задача выделить сферу деятельности компаний, оперирующих данными (data companies – компании, «основной деятельностью которых является производство и поставка продуктов, услуг и технологий, связанных с цифровыми данными») [4]. Для этого были определены более широкие коды классификатора видов экономической деятельности, в рамках которых выделялась совокупность компаний, оперирующих данными – в рамках проекта такими видами экономической деятельности были классы J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая) [4].

Выделение индустрии работы с большими данными и измерение показателей, ее характеризующих, проводится на базе выборочного представительного опроса организаций цифрового сектора экономики. В качестве критериев выделения компаний индустрии работы с большими данными и измерения показателей предлагается использовать предложенную в таблице 1 классификацию видов деятельности, составляющих индустрию работы с большими данными. Кроме того, в проводимом исследовании будет учитываться глобальное руководство и таксономия расходов на большие данные и аналитику, которое разработано и поддерживается в актуальном состоянии компанией IDC, многие годы занимающейся измерением глобального рынка аналитики больших данных [5]. Таксономия IDC технологий, решений и услуг работы с большими данными позволит выделить производителей и объемы выручки по продуктовым позициям в сфере работы с большими данными и проводить международные сопоставления. Наконец, для идентификации компаний индустрии работы с большими данными будет использоваться Единый реестр российских программ для электронных вычислительных машин и баз данных [6], для которого разработан более детальный классификатор, выделяющий программное обеспечение для работы с большими данными (см. в частности, раздел «10. Средства обработки и визуализации данных» и «11. Средства анализа данных»).

2 Обзор литературы

2.1. Обзор подходов, концептуальных схем и показателей, разработанных международными организациями и аналитическими компаниями для мониторинга сферы производства и рынка цифровых технологий

Как уже отмечалось, действующие классификаторы видов экономической деятельности, товаров и услуг не позволяют статистически выделить индустрию работы с большими данными и измерять на этой основе ее показатели.

Вместе с тем, международными организациями проделана большая работа по разработке и развитию стандартов измерения более широкого сектора экономики, частью которого является индустрия работы с большими данными, – информационной индустрии (цифрового сектора экономики, включающего ИКТ-сектор и сектор контента и СМИ). Опыт этого измерения, прежде всего в части предложенных показателей, использован в рамках данной работы.

Впервые определение и группировку видов экономической деятельности, относящихся к ИКТ-сектору, предложила ОЭСР в 1998 г. Впоследствии группировка дорабатывалась и адаптировалась к новой версии международного классификатора видов экономической деятельности, к ней была добавлена группировка сектора контента и СМИ [1]. В 2015 г. приказом Минкомсвязи России [2] были утверждены аналогичные группировки на основе ОКВЭД-2, согласованные с группировками ОЭСР.

Сектор ИКТ в рамках этого подхода определяется как совокупность видов экономической деятельности, связанных с производством товаров и услуг, предназначенных для выполнения функции обработки информации и коммуникации с использованием электронных средств, включая передачу и отображение информации. На основе этого определения был составлен перечень соответствующих кодов международного классификатора экономической деятельности. В состав сектора ИКТ вошли производство и оптовая торговля ИКТ-товарами (в частности,

компьютерного и коммуникационного оборудования), а также оказание ИКТ-услуг (телекоммуникационных услуг и услуг на основе информационных технологий – ИТ-услуг).

Сектор контента и СМИ определяется как совокупность видов экономической деятельности, осуществляющих производство, издание и распространение контента (информации, продукции культурного назначения и продукции, предназначенной для развлекательных целей).

Также для измерения продукции цифрового сектора, ее экспорта и импорта ОЭСР были разработаны классификаторы товаров (первый вариант в 2003 г.) и услуг (2006 г.) ИКТ. Аналогичная работа была проделана для сектора контента и СМИ. Соответствующие классификаторы представляют собой собирательные группировки общих классификаторов товаров и услуг [1]. В 2015 г. приказом Минкомсвязи России были утверждены аналогичные группировки [2].

Для измерения цифрового сектора международные организации предложили и используют целый ряд показателей. Их набор ни одна из них не организует в виде концептуальной схемы, но их можно сгруппировать в несколько категорий: показатели масштабов, конкурентоспособности и потенциала развития информационной индустрии.

Сама ОЭСР для измерения ИКТ-сектора и сектора контента и СМИ использует несколько показателей. Два из них характеризуют размер (масштабы) сектора – (1) доля сектора в ВВП (или в добавленной стоимости, созданной в предпринимательском секторе); (2) доля занятых в секторе от всех занятых в экономике или в предпринимательском секторе [7, 8]. Международная торговля ИКТ-товарами и услугами представлена показателями, которые можно использовать для построения показателей конкурентоспособности – доли ИКТ-продукции в общем экспорте товаров и услуг [8,9].

Аналогичные показатели для ИКТ-сектора рекомендует международное Партнерство по измерению ИКТ для развития, добавляя в своем перечне ключевых показателей ИКТ к показателям международной торговли импорт товаров ИКТ в процентах от совокупного объема импорта [10].

Используемый Евростатом набор показателей ИКТ-сектора несколько шире. Помимо показателей размера сектора используется показатель темпов роста сектора, а также два показателя, характеризующих инновационный потенциал сектора ИКТ – (1) доля расходов на исследования и разработки сектора ИКТ в общих расходах на исследования и разработки и (2) доля персонала, занятого НИР в секторе от всего исследовательского персонала страны [11]. К характеристике потенциала развития ИКТ-сектора можно отнести показатель числа молодых быстрорастущих предприятий сектора (gazelles) [12].

В России НИУ ВШЭ совместно с Росстатом и Минцифры России с 2009 г. выпускает ежегодный статистический сборник «Индикаторы цифровой экономики» (до 2016 г. под названием «Индикаторы информационного общества»), посвященный мониторингу производства и использования цифровых технологий [13]. Один из разделов сборника посвящен мониторингу предприятий ИКТ-сектора и сектора контента и СМИ. Среди показателей – количество предприятий ИКТ-сектора, численность занятых, валовая добавленная стоимость и инвестиции в основной капитал. На основе выборочного обследования предприятий в сборнике также оценивается деловая активность организаций, оказывающих услуги в сфере ИКТ. Под деловой активностью понимаются конкурентоспособность, экономическое положение организации, объем продаж разработанного ПО на внутреннем рынке и за рубежом, спрос на услуги организации и т.д. Отдельный блок показателей характеризует структуру затрат на инновационную деятельность организаций ИКТ-сектора.

Возвращаясь к проблеме выделения предприятий индустрии работы с большими данными из предприятий цифрового сектора экономики необходимо обратить внимание на решение аналогичной проблемы, которое было предложено в рамках уже упоминавшегося выше проекта Европейской комиссии «Европейский рынок данных» [4]. Помимо прочего перечень разработанных в рамках проекта показателей включал в себя число компаний, оперирующих данными, и их выручку. Под компаниями, оперирующими данными, в исследовании понимались организации (стартапы, малые, средние предприятия или подразделения крупных компаний) которые занимаются разработкой и внедрением новых технологий и решений для работы с большими данными или усовершенствованием и обновлением старых. Таким образом, согласно приведенному выше определению, индустрия работы с большими данными включает в себя как «чистых игроков», чьим основным бизнесом являются продукты и услуги для работы с большими

данными, так и «смешанных игроков», которые предлагают как традиционные, так и основанные на работе с большими данными решения.

Оценка числа компаний, оперирующих данными и их выручки производилась в проекте Европейской комиссии с помощью выборочного обследования компаний, чья экономическая деятельность относится к классам J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая) согласно статистической классификации экономической деятельности в Европейском сообществе NACE Rev. 2 [14]. Выборка компаний состояла примерно из 1500 организаций. В нее вошли две группы компаний (по численности сотрудников): менее 250 сотрудников и 250 и более сотрудников. При этом опрос проводился среди независимых компаний, а не местных филиалов крупных транснациональных корпораций. Выборочное обследование компаний индустрии работы с большими данными было дополнено проведенным отдельно обследованием стартапов в данной отрасли. На основе полученных данных делалась оценка числа и выручки компаний, работающих в индустрии больших данных, в странах ЕС (доля компаний, производящих и занимающихся внедрением новых технологий, в выборке экстраполировалась на все компании, чья деятельность относится к классам J или M). Число компаний индустрии работы с большими данными в тех странах ЕС, где выборочного обследования не проводилось, также оценивалось по специальной методике с помощью экстраполяции данных, полученных в других странах. По аналогичной методике оценивалась выручка европейских компаний данного сектора экономики.

Таким образом, сегодня уже существуют достаточно проработанные подходы и наборы показателей, которые успешно применяются для мониторинга сферы производства и рынка цифровых технологий в России и мире. Данные показатели можно адаптировать для концептуальной схемы мониторинга сектора экономики, производящего технологии, решения и услуги работы с большими данными.

2.2. Обзор исследований и аналитических докладов, посвященных анализу индустрии данных и индустрии работы с большими данными

Несмотря на бурное развитие и использование технологий работы с большими данными в различных сферах человеческой деятельности, научная литература пока представлена сравнительно небольшим числом работ, посвященных измерению, оценке и анализу сектора экономики, разрабатывающего данный класс цифровых технологий, решений и сервисов на их основе. Это связано в том числе с обозначенными выше проблемами статистического учета предприятий-производителей и поставщиков продуктов и услуг для работы с большими данными, когда существующие классификации экономической деятельности оказываются слишком «грубыми» для данной сферы деятельности.

Так хотя поиск по специально разработанному поисковому образу, включающему все основные термины, связанные с технологиями работы с большими данными в библиографических базах данных платформы Web of Science [15], позволяет найти 108067 публикаций за 2016-2020 гг., абсолютное большинство из них посвящены анализу новых технологий, возможностей их применения и влиянию, которое они могут оказать на те или иные сферы деятельности. Поэтому для поиска релевантных статей необходимо сузить поисковый образ до двух терминов «индустрия работы с большими данными» (“big data industry”) и «компании, оперирующие большими данными» (“big data compan*”). Поиск проводился по полю «Тема», включающему названия, ключевые слова и аннотации статей. В результате по данному поисковому запросу было найдено 84 публикации, изданные в 2016-2020 гг. Далее содержание найденных работ было изучено с целью отбора релевантных статей. В общей сложности было отобрано 10 наиболее релевантных публикаций, посвященных анализу индустрии работы с большими данными, остальные 74 статьи были исключены из анализа, поскольку они не имели прямого отношения к исследуемой проблеме.

Значительная часть проанализированной научной литературы, посвященной индустрии работы с большими данными, носит описательный характер: в статьях рассматриваются основные технологии и решения для работы с большими данными существующие сегодня на рынке, их влияние на рынок и особенности ценообразования [16; 17]; влияние на индустрию больших данных отдельных технологий [18] однако не делается попытки измерить развитие данного сектора экономики. Некоторые исследователи делали попытки выявить и оценить факторы, влияющие на развитие индустрии работы с большими данными в отдельных регионах [19], а также взаимосвязь с «зеленой» экономикой [20].

Несколько иной подход выбрали авторы статьи [21], попробовав оценить развитие индустрии работы с большими данными в Китае по пяти категориям: 1) промышленный масштаб и уровень использования технологий работы с большими данными; 2) развитие предприятий-разработчиков и поставщиков технологий работы с большими данными; 3) технический уровень решений в области работы с большими данными; 4) особенности применения технологий работы с большими данными в промышленности Китая; 5) политическая среда развития индустрии работы с большими данными в стране. На основе анализа по перечисленным предметным областям авторы выявляют основные барьеры развития индустрии работы с большими данными и формируют рекомендации по их преодолению.

Схожее исследование по выявлению препятствий для развития индустрии работы с большими данными и разработке рекомендаций для корректировки политики в данной области было проведено в Корее [22]. Большинство подобных исследований строилось на обследовании определенной выборки предприятий-разработчиков и поставщиков технологий, услуг и решений работы с большими данными (как правило, это были самые крупные игроки на рынке), а выводы, сделанные на основе обследования данных выборок, распространялись на весь рынок [23; 24]. Из-за отсутствия надежных количественных статистических данных большинство исследований посвященных изучению индустрии работы с большими данными опираются на качественные методы, такие как интервью экспертов и представителей компаний, фокус-группы, SWOT-анализ и т.д. Исходя из результатов интервью и опросов, исследователи пытаются выявить основные направления развития индустрии работы с большими данными и разработать стратегии развития данного сектора экономики [25].

Таким образом, сегодня в мире не существует надежного методологического инструментария для оценки уровня развития сектора экономики, производящего технологии, решения и услуги, связанные с большими данными, а существующая статистика и действующие классификаторы не позволяют «вычленивать» индустрию работы с большими данными из общего сектора ИКТ. Поэтому большинство исследований индустрии работы с большими данными, посвящены либо отдельным предприятиям и технологиям (case-study), либо основаны на выборочных обследованиях сравнительно небольшого количества предприятий и экспертных опросах.

В описанных выше рекомендациях международных организаций и практике статистического наблюдения за развитием более широкого цифрового сектора экономики (или его части – ИКТ-сектора) сложились подход и система показателей, которые, как отмечалось выше, можно сгруппировать в три предметные области мониторинга: масштаб (доля сектора в ВВП и рабочей силы в занятом населении), конкурентоспособность (доля в общем экспорте и импорте, соотношение экспорта и импорта) и инновационная активность предприятий сектора, которая связана с потенциалом его развития (расходы предприятий на НИОКР и др.) [7-13].

3 Концептуальная схема и показатели мониторинга

3.1 Концептуальная схема предметной области

Для мониторинга индустрии работы с большими данными предложенная в предыдущем разделе группировка показателей была взята за основу концептуальной схемы, учитывая ее комплексный характер и многолетний опыт использования входящих в эту группировку показателей для измерения цифрового сектора. Вместе с тем состав показателей был доработан с учетом других систем мониторинга, потребностей аудитории мониторинга и дополнительных возможностей, которые предоставляет использование в рамках мониторинга развития и использования технологий работы с большими данными целого спектра методов измерения (в частности, экспертных оценок и опроса предприятий).

В таблице 2 представлена концептуальная схема и набор показателей для мониторинга индустрии работы с большими данными, которые будут рассчитываться для предприятий, выделяемых в соответствии с подходом и классификацией, представленными в предыдущем разделе.

Таблица 2. Концептуальная схема мониторинга производства технологий работы с большими данными

Масштабы производства технологий работы с большими данными	Конкурентоспособность производства технологий работы с большими данными	Потенциал производства технологий работы с большими данными
Доля индустрии работы с большими данными в общем объеме выручки цифрового сектора экономики	Доля российских поставщиков в общем объеме мирового экспорта технологий, товаров и услуг работы с большими данными	Доля российских стартапов в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными
Доля занятых в индустрии работы с большими данными, от общего числа занятых в цифровом секторе экономики	Доля российских производителей на мировом рынке технологий, товаров и услуг работы с большими данными	Соотношение отечественного и мирового уровней готовности технологий работы с большими данными
	Доля российских производителей на отечественном рынке технологий, товаров и услуг работы с большими данными	Доля проектов, связанных с производством технологий работы с большими данными, в общем числе проектов, поддержанных институтами развития
	Соотношение экспорта и импорта технологий, решений и услуг работы с большими данными	

3.2 Показатели масштабов производства

Как отмечалось в обзоре литературы, для характеристики масштабов отрасли экономики обычно используются несколько ключевых показателей, таких как доля этой отрасли в валовом внутреннем продукте, а также доля работников отрасли в общем числе занятого населения (международное Партнерство по измерению ИКТ для развития, добавляет к ним показатели международной торговли товарами ИКТ, такие как доля импорта ИКТ-товаров от общего объема импорта страны и доля экспорта ИКТ-товаров в общем объеме товарного экспорте страны [10]).

Для мониторинга индустрии работы с большими данными используется более широкая концептуальная схема, показатели международной торговли товарами и услугами работы с большими данными рассматриваются в контексте оценки конкурентоспособности индустрии. Что касается показателей, характеризующих масштаб индустрии работы с большими данными, предлагается их использовать в несколько модифицированном виде и рассчитывать долевые показатели добавленной стоимости и числа занятых не от показателей экономики в целом, а по отношению к цифровому сектору.

Выбор в качестве базы для расчета показателей цифрового сектора экономики, а не всей экономики (или ее предпринимательского сектора) связан с несколькими соображениями. Сам объем индустрии работы с большими данными пока невелик, значение показателей будет составлять доли процента от ВВП и с такими значениями, во-первых, трудно оперировать, а во-вторых, в таком варианте значения показателей будут чувствительны к изменению общей структуры экономики (например, повышение цен на энергетические ресурсы повышает долю ТЭК в экономике за счет других отраслей). Рассмотрение индустрии работы с большими данными в контексте общих показателей цифрового сектора позволит оценить относительную значимость этой индустрии во всем комплексе экономической деятельности, связанной с цифровыми технологиями, при этом всегда возможен простой пересчет показателей индустрии по отношению

к общеэкономическим показателям. Такой подход, отметим, принят в упомянутом выше исследовании «Европейский рынок данных» [4] – для характеристики «индустрии данных» используется, помимо общего числа компаний, оперирующих данными, и объема их выручки, показатели доли этих компаний не по отношению к экономике в целом, а в общем числе и выручке компаний, входящих в классы J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая).

В итоге, для характеристики масштабов индустрии работы с большими данными в данной работе будут использоваться следующие показатели:

- (МИБД-01-01) *Доля индустрии работы с большими данными в общем объеме выручки цифрового сектора экономики*

Использование предложенного показателя позволяет оценить масштаб индустрии работы с большими данными относительно всего цифрового сектора, а также его динамику. При этом абсолютные объемы производства товаров и услуг для работы с большими данными (выраженные, например, в рублевом эквиваленте), так же, как и динамика этих значений от года к году, могут быть вычислены напрямую на основе абсолютных объемов производства товаров и услуг цифрового сектора экономики.

- (МИБД-01-02) *Доля занятых в индустрии работы с большими данными, от общего числа занятых в цифровом секторе экономики*

Предложенный показатель демонстрирует социально-экономическую значимость индустрии работы с большими данными относительно всего цифрового сектора экономики; абсолютное значение и динамика от года к году также могут быть вычислены.

Так как официальные статистические данные не выделяют производство товаров и услуг для работы с большими данными из цифрового производства, основным инструментом для оценки указанных показателей является специализированное исследование, включающее представительные опросы предприятий цифрового сектора (см. выше).

3.3 Показатели конкурентоспособности производства

Конкурентоспособность отечественного производства технологий работы с большими данными характеризуется их востребованностью на отечественном и мировом рынках. Оценка конкурентоспособности проводится на основе измерений соотношения экспорта и импорта, доли российских производителей на рынках технологий работы с большими данными. Для мониторинга конкурентоспособности используется четыре показателя.

- (КИБД-02-01) *Доля российских поставщиков в общем объеме мирового экспорта технологий, решений и услуг работы с большими данными*

Показатель демонстрирует «вес» России среди мировых экспортеров технологий, товаров и услуг работы с большими данными. Более высокое значение показателей говорит о более высокой конкурентоспособности отечественных производителей.

- (КИБД-02-02) *Доля российских производителей на мировом рынке технологий, решений и услуг работы с большими данными*

Показатель демонстрирует востребованность продукции российских производителей на мировом рынке, а динамика показателя год к году демонстрирует изменение (рост или снижение) конкурентоспособности российских производителей. Более высокое значение показателя говорит о более высокой конкурентоспособности российских производителей.

- (КИБД-02-03) *Доля отечественных поставщиков на российском рынке технологий, решений и услуг работы с большими данными*

Показатель демонстрирует востребованность российских производителей на отечественном рынке, конкурентоспособность на внутреннем рынке товаров и услуг индустрии работы с большими данными по сравнению с импортными. Динамика показателя год к году демонстрирует изменение (рост или снижение) конкурентоспособности российских производителей на внутреннем рынке, более высокое значение показателя говорит о более высокой конкурентоспособности российских поставщиков.

- (КИБД-02-04) *Соотношение экспорта и импорта технологий, решений и услуг работы с большими данными*

Показатель демонстрирует, является ли Россия нетто-экспортером или нетто-импортером технологий работы с большими данными. Более высокое значение показателя говорит о способности российской индустрии работы с большими данными удовлетворять потребности отечественной экономики в технологиях, решениях и услугах работы с большими данными.

Предложенная система показателей позволяет с разных сторон оценить востребованность, а через нее – конкурентоспособность отечественного производства технологий работы с большими данными. При этом, оценивается конкурентоспособность как на отечественном, так и на мировом рынке.

Инструментом оценки показателей КИБД-02-01 – КИБД-02-04 является специализированное исследование, включающее обследование предприятий цифрового сектора экономики.

3.4 Показатели потенциала производства

Оценивая состояние индустрии работы с большими данными, важно характеризовать и потенциал ее развития – наличие условий для роста и конкурентоспособности отечественных производителей технологий работы с большими данными. Потенциал развития любой высокотехнологичной отрасли экономики обычно связывают с развитием существующих и разработкой новых технологий. В случае индустрии работы с большими данными потенциал оценивается тремя показателями – количеством новых стартапов и уровнем готовности наиболее зрелых технологий работы с большими данными, разрабатываемых отечественными производителями, относительно аналогичных технологий мировых производителей, а также долей проектов, связанных с производством технологий работы с большими данными, в общем числе проектов, поддержанных институтами развития.

- (ПИБД-03-01) *Доля российских стартапов в общем числе стартапов, связанных с технологиями работы с большими данными.*

Количество новых стартапов (относительно мирового их числа) демонстрирует активность отечественных разработчиков на рынке больших данных, которая, соответственно, потенциально может привести к разворачиванию производства новых технологий и решений для работы с большими данными, востребованных на отечественном и мировом рынках, к формированию новых сегментов рынков. Большее значение показателя свидетельствует о более высоком потенциале отечественной индустрии работы с большими данными.

Исходные данные для оценки показателя числа стартапов берутся из международной базы данных Crunchbase [26].

- (ПИБД-03-02) *Соотношение отечественного и мирового уровня готовности технологий работы с большими данными*

Уровень готовности технологии (Technology Readiness Level, TRL) и уровень готовности производства (Manufacturing Readiness Level, MRL) [27] являются общепринятыми методами измерения зрелости технологий и их производства, применимыми и для технологий работы с большими данными. Показатель будет измеряться методом экспертных опросов. Предполагается, что перед оценкой этого показателя экспертами будет выделяться перечень наиболее перспективных и востребованных технологий работы с большими данными, по каждой из которых будет выполнено соответствующее сопоставление отечественного и мирового уровней. Таким образом будет обеспечена актуальность оценки и сопоставимость с мировым уровнем.

- (ПИБД-03-03) *Доля проектов, связанных с производством технологий работы с большими данными, в общем числе проектов ИКТ, поддержанных институтами развития*

Показатель поддержки проектов в сфере производства технологий работы с большими данными отечественными институтами развития, работающими с проектами, находящимися на высоком уровне технологической и производственной готовности, отражает не только активность отечественных институтов развития в поддержке проектов, связанных с большими данными, но и является важной характеристикой потенциала развития индустрии работы с большими данными, отражая наличие в стране достаточного числа технологических инициатив, развивающих перспективные технологии.

4 Методология и результаты

Полномасштабная реализация разработанной методологии предполагает проведение комплексного обследования рынка и производителей технологий, товаров и услуг работы с большими данными, которое выходит за рамки возможностей пилотной реализации. В данном разделе представлены те показатели, измерение которых было возможно методом кабинетных исследований.

4.1 Методы измерения

Одним из показателей потенциала производства технологий работы с большими данными в России является количество новых российских стартапов в данной области и их доля от стартапов созданных во всем мире. В качестве источника информации о количестве созданных стартапов в России и мире в нашем исследовании используется база данных Crunchbase [26] – одна из крупнейших в мире баз данных, содержащая деловую информацию о частных и государственных компаниях. Данные Crunchbase поступают из различных каналов – от пользователей базы, представителей компаний, общедоступных источников или различных партнеров по сбору данных (например, инвесторских компаний). Хотя Crunchbase не претендует на всеобъемлющий охват своих данных и в некоторых случаях информация по одним отраслям или регионам может быть менее полной, чем по другим, на данный момент большинством ученых-экономистов эта база признается в качестве наиболее репрезентативной из существующих и активно используется в исследованиях посвященных стартапам, развитию инновационного бизнеса и цифровой экономики, в том числе при проведении международных сопоставлений [28; 29]. Данные Crunchbase также успешно применялись при исследовании российских стартапов [30].

В научной литературе нет консенсуса относительно того, какие компании можно считать стартапами, а какие нет. Как правило под стартапами понимают молодые, нацеленные на рост компании. Разные исследователи считают, что стартапами можно считать компании возрастом до одного, трех, шести или даже десяти лет [31]. Для целей данного мониторинга стартапами будут считаться компании, связанные с технологиями, товарами и услугами работы с большими данными, не старше 3 лет (т.е., например, если говорить про 2020 г., то стартапами мы считаем компании, зарегистрированные с 2018 по 2020 г. включительно).

Для отбора российских компаний, связанных с технологиями работы с большими данными, был сформирован поисковый образ, включающий 25 отраслей (по классификации Crunchbase), непосредственно относящихся к разработке и внедрению технологий работы с большими данными (таблица 3). Стоит отметить, что в классификации Crunchbase есть как достаточно широкие технологические отрасли (например, «Искусственный интеллект»), так и более узкие, привязанные к конкретным видам технологий (например, «Обработка естественного языка»). Учитывая, что компании в базе данных Crunchbase могут относиться одновременно к нескольким отраслям (например, к доставке еды и большим данным), разработанный нами поисковый образ позволяет создать по формальным признакам максимально репрезентативную выборку компаний, занимающихся разработкой и внедрением технологий, услуг и решений для работы с большими данными в различных отраслях экономики.

Таблица 3. Список отраслей по классификации Crunchbase, использованный для поиска компаний, занимающихся разработкой и (или) внедрением технологий, решений и услуг для работы с большими данными

№ п/п	Название отрасли
1	Аналитика (Analytics)
2	Искусственный интеллект (Artificial Intelligence)
3	Автономные транспортные средства (Autonomous Vehicles)
4	Большие данные (Big Data)
5	Бизнес-аналитика (Business Intelligence)
6	Облачные сервисы для работы с данными (Cloud Data Services)
7	Компьютерное зрение (Computer Vision)

№ п/п	Название отрасли
8	Центр обработки данных (Data Center)
9	Автоматизация центра обработки данных (Data Center Automation)
10	Интеграция данных (Data Integration)
11	Интеллектуальный анализ данных (Data Mining)
12	Хранение данных (Data Storage)
13	Визуализация данных (Data Visualization)
14	Распознавание лиц (Facial Recognition)
15	Распознавание изображений (Image Recognition)
16	Интеллектуальные системы (Intelligent Systems)
17	Машинное обучение (Machine Learning)
18	Обработка естественного языка (Natural Language Processing)
19	Предиктивная аналитика (Predictive Analytics)
20	Поисковая система (Search Engine)
21	Семантический поиск (Semantic Search)
22	Распознавание речи (Speech Recognition)
23	Интеллектуальный анализ текстов (Text Analytics)
24	Виртуальный помощник (Virtual Assistant)
25	Визуальный поиск (Visual Search)

4.2 Результаты

По данным Crunchbase, на 2020 г. в мире существовало 9619 стартапов (т.е. компаний, созданных в 2018–2020 гг.), занимающихся разработкой и внедрением технологий, решений и услуг для работы с большими данными, из них российских компаний – 50 (0,52%). Интересной тенденцией последних трех лет является постепенное сокращение количества стартапов. Так, с 2017 г. число стартапов сократилось более чем на треть: с 16 тыс. до менее 10 тыс. во всем мире, в России – с 81 до 51. Особенно резкое сокращение наблюдается в 2020 г., что, по-видимому, связано с глобальной пандемией коронавируса, а также с тем, что данные базы Crunchbase могут пополняться с некоторым запозданием.

Абсолютным лидером в мире по количеству интересующих стартапов являются США. В то же время их доля в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными, в последние годы стабильно падает. Так, в 2017 г. в Crunchbase была информация о 5610 молодых американских компаний, что составляло 37,4% от всех зарегистрированных в базе стартапов, работающих с большими данными. А уже в 2020 г. эта доля сократилась до 33% (3179 компаний). Еще более резкое падение числа новых компаний наблюдается в Китае: доля китайских компаний в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными упала с 6,1% (909 стартапов) в 2017 г. до 1,9% (181 стартап). Рост по данному показателю среди стран-лидеров в последние годы наблюдался только в Германии (2,6% до 3,2%) и Японии (с 1,8% до 2,3%). Данные о динамике доли стартапов стран, являющихся технологическими лидерами, в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными, за последние пять лет представлены на рисунке 1.

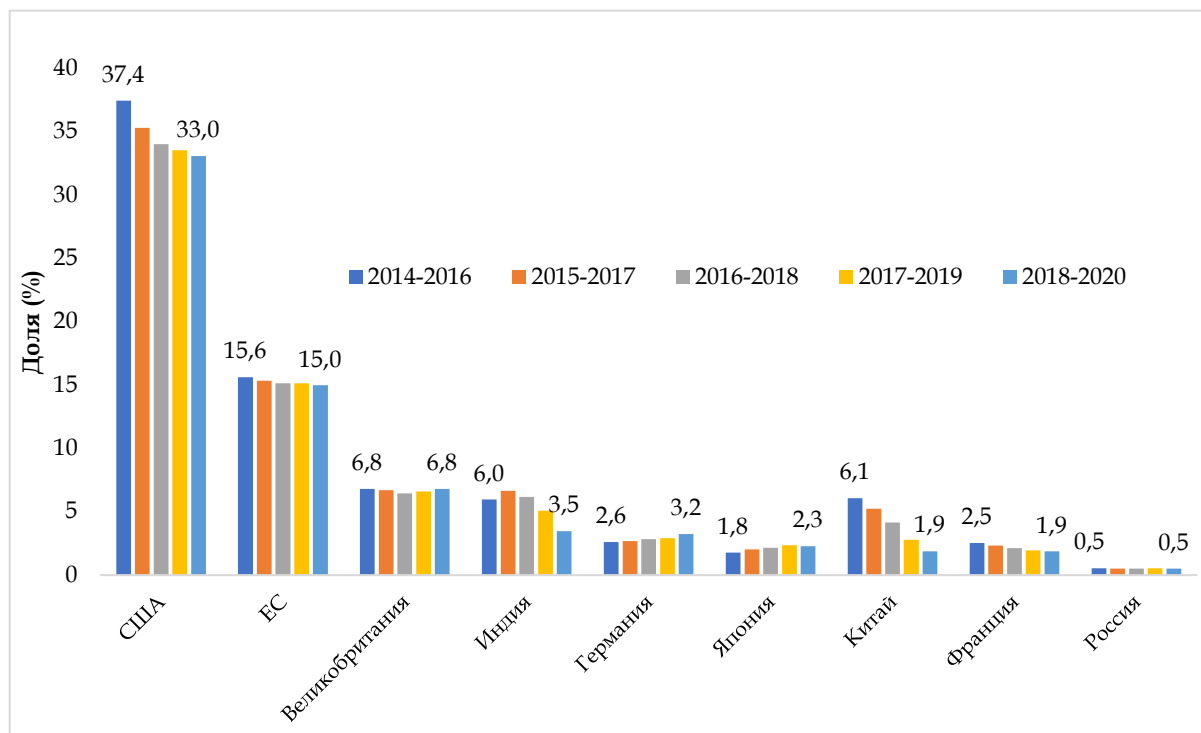


Рисунок 1. Динамика изменения доли новых стартапов различных стран в общем числе мировых стартапов, связанных с технологиями, товарами и услугами работы с большими данными (2014–2020 гг.)
Источник: составлено авторами на основе данных Crunchbase

Как видно из графика, Россия серьезно отстает от большинства развитых и многих развивающихся стран по значению показателя, характеризующего долю российских стартапов в общем числе мировых стартапов, связанных с технологиями, товарами и услугами работы с большими данными. Это свидетельствует о сравнительно низком уровне развития индустрии больших данных в России. В то же время, в отличие от большинства стран-лидеров в данной области, доля российских компаний среди мировых стартапов стабильно держится на уровне 0,5%, не снижаясь последние пять лет.

Заключение

В условиях отсутствия статистических инструментов для измерения индустрии работы с большими данными (необходимой для этого детализации классификаторов видов экономической деятельности, товаров и услуг работы с большими данными) для мониторинга этого важного сектора отечественной экономики был разработан представленный в статье подход, включающий:

- определение и классификацию видов деятельности, которые относятся к индустрии работы с большими данными (основана на модели работы с большими данными, представленной в стандарте эталонной архитектуры больших данных Международной организации по стандартизации);
- концептуальную схему и систему показателей для мониторинга;
- методы измерения показателей, основанные на опросе предприятий цифрового сектора и потребителей технологий, решений и услуг работы с большими данными.

Одним из основных направлений дальнейшего развития методологии мониторинга является разработка таксономии технологий, товаров (решений) и услуг работы с большими данными, а также ее тестирование в ходе проведения опроса предприятий цифрового сектора и потребителей технологий, решений и услуг работы с большими данными.

Измерение одного из показателей методологии показало, что на такой авторитетной международной площадке, как Crunchbase российские компании и стартапы пока составляют небольшую долю. Одним из возможных факторов низкого значения показателя среди прочего может быть недостаточно полное освещение данных по российским компаниям в базе Crunchbase по сравнению с компаниями из Северной Америки или Западной Европы. Учитывая, что

источниками данных Crunchbase являются в том числе и представители компаний, для исправления ситуации можно рекомендовать в обязательном порядке всем российским стартапам вносить информацию в данную базу.

Разработанный подход к мониторингу индустрии работы с большими данными может быть использован для создания систем мониторинга развития, производства и использования «сквозных» цифровых технологий – для них возникают аналогичные проблемы при определении производителей решений и услуг, основанных на этих технологиях.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. OECD. OECD Guide to Measuring the Information Society 2011. Paris: OECD Publishing, 2011. <https://doi.org/10.1787/9789264113541-en>
2. Приказ Минкомсвязи России от 7 декабря 2015 года № 515 «Об утверждении собирательных классификационных группировок "Сектор информационно-коммуникационных технологий" и "Сектор контента и средств массовой информации"». URL: <http://docs.cntd.ru/document/420327966> (дата обращения 01.06.2021).
3. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture. URL: <https://www.iso.org/standard/71277.html> (дата обращения 01.06.2021).
4. The European Data Market Study: Final Report. By IDC and Open Evidence. 2017. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirects=0&d=1 (дата обращения 01.06.2021)
5. Goepfert J. et al. IDC's Worldwide Big Data and Analytics Spending Guide Taxonomy, IDC. URL: <https://www.idc.com/getdoc.jsp?containerId=US47485920> (дата обращения 01.06.2021)
6. Единый реестр российских программ для электронных вычислительных машин и баз данных. URL: <https://reestr.digital.gov.ru/> (дата обращения 01.06.2021)
7. OECD. Size of the ICT Sector. 2013. URL: <https://www.oecd-ilibrary.org/docserver/factbook-2011-72-en.pdf?expires=1632472701&id=id&accname=guest&checksum=8C61EC5E182618032918749217A03978> (дата обращения 01.06.2021)
8. OECD. Measuring the Digital Transformation. A Roadmap for the Future. Paris; OECD Publishing, 2019. <https://doi.org/10.1787/9789264311992-en> (дата обращения 01.06.2021)
9. OECD Key ICT Indicators. URL: <https://www.oecd.org/sti/ieconomy/oecdkeyictindicators.htm> (дата обращения 01.06.2021)
10. Core list of indicators. March 2016 version. Partnership on Measuring ICT for Development. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf (дата обращения 01.06.2021).
11. ICT sector (isoc_se). Reference Metadata in Euro SDMX Metadata Structure (ESMS). Eurostat, the statistical office of the European Union. URL: https://ec.europa.eu/eurostat/cache/metadata/en/isoc_se_esms.htm (дата обращения 01.06.2021).

12. Eurostat. Browse statistics by theme. Digital economy and society. Database. ICT sector. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (дата обращения 01.06.2021)
13. Индикаторы цифровой экономики: 2020: статистический сборник / Г. И. Абдрахманова, К. О. Вишнеvский, Л. М. Гохберг и др.; Нац. исслед. ун-т «Высшая школа экономики». – М.: НИУ ВШЭ, 2020. URL: <https://issek.hse.ru/mirror/pubs/share/387609461.PDF> (дата обращения 01.06.2021).
14. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European. - Luxembourg: Office for Official Publications of the European Communities, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (дата обращения 01.06.2021)
15. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
16. Rojas-Torres D., Kshetri N. Big data solutions for micro-, small-, and medium-sized enterprises in developing countries // IT Professional. 2019. Vol. 21. №. 5. P. 67-70. <https://doi.org/10.1109/MITP.2019.2932236>
17. Liang F. et al. A survey on big data market: Pricing, trading and protection // Ieee Access. 2018. Vol. 6. P. 15132–15154. <https://doi.org/10.1109/ACCESS.2018.2806881>
18. Ahmad M., Rashid T., Mishra D. K. Impact of wireless technology on future of big-data industry // 2014 Conference on IT in Business, Industry and Government (CSIBIG). IEEE, 2014. P. 1–4, <https://doi.org/10.1109/CSIBIG.2014.7056931>
19. Li C., Pan K. Research on financing efficiency of big data industry based on three stage DEA-Taking Guizhou Province as an example // Advances in Social Science, Education and Humanities Research (ASSEHR), 2018. Vol. 130. P. 439–442. <https://doi.org/10.2991/ifmeita-17.2018.74>
20. Zhang N., Zhang H., Zhang M. The Harmonious Development of Big Data Industry and Ecological Civilization Construction in Guizhou // 2018 6th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM 2018). Atlantis Press, 2019. P. 192–198. <https://doi.org/10.2991/ieesasm-18.2019.36>
21. Liu Y. et al. An overview of big data industry in China // China Communications. 2014. Vol. 11. № 12. P. 1–10. <https://doi.org/10.1109/CC.2014.7019834>
22. Kwon T. H., Kwak J. H., Kim K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea // Technological Forecasting and Social Change. – 2015. – Т. 96. – С. 144–152. <https://doi.org/10.1016/j.techfore.2015.03.017>
23. Caputo F. et al. Over the mask of innovation management in the world of Big Data // Journal of Business Research. 2020. Vol. 119. P. 330–338. <https://doi.org/10.1016/j.jbusres.2019.03.040>
24. Tang C. The data industry: The business and economics of information and big data. Hoboken: John Wiley & Sons, 2016, 195 p.
25. Weng W. H., Lin W. T. Development Trends and Strategy Planning in Big Data Industry // Contemporary Management Research. 2014. Vol. 10. № 3. P. 203–214. <https://doi.org/10.7903/cmr.12288>
26. Crunchbase. 2021. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirects=0&d=1 (дата обращения 01.06.2021).
27. ГОСТ Р 58048–2017 Трансфер технологий. Методические указания по оценке уровня зрелости технологий; ГОСТ Р 57194.1–2016 Трансфер технологий. Общие положения. URL: <https://docs.cntd.ru/document/1200158331> (дата обращения 01.06.2021).
28. Dalle J. M., Besten M., Menon C. Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers, No. 2017/08. Paris: OECD Publishing, 2017. 29 p. <https://doi.org/10.1787/6c418d60-en>
29. Kézai P. K., Fischer S., Lados M. Smart economy and startup enterprises in the Visegrád Countries – A comparative analysis based on the crunchbase database // Smart Cities. 2020. Vol. 3. №. 4. P. 1477–1494. <https://doi.org/10.3390/smartcities3040070>
30. Akhmetova J. A. The Impact of Qualitative Determinants on Startup Investments in Russia // Наука, технологии и бизнес: сборник материалов II межвузовской заочной конференции аспирантов, соискателей и молодых ученых, Москва, 28–29 апреля 2020 года. Москва:

- Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет), 2020. С. 35–41.
31. Иншаков М. О., Орлова А. А. Инновационные стартапы в России: проблемы создания и маркетингового продвижения // Вестник Волгоградского государственного университета. Серия 3: Экономика. Экология. 2014. №. 1. С. 66–76.

BIG DATA INDUSTRY

Malakhov, Vadim Aleksandrovich

Candidate of historical sciences

Vavilov Institute for the History of Science and Technology of the Russian Academy of Sciences, Center for the history of science organization and scientific research, senior researcher

Moscow, Russia

yasonbh@mail.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor

Moscow, Russia

yuri.hohlov@iis.ru

Shaposhnik, Sergei Borisovich

Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher

Petrozavodsk, Russian Federation

sergei.shaposhnik@gmail.com

Abstract

The article proposes an approach, a conceptual framework and indicators for monitoring the big data industry, which is understood as a set of companies producing technologies, solutions and services in the field of big data. The conceptual framework for monitoring is based on the role model of big data, as presented in the International Organization for Standardization (ISO) standard for big data reference architecture. The monitoring framework includes 9 indicators characterizing the scale, competitiveness, and potential of the big data industry. Measurement of the indicator related to the number of startups in the big data industry, based on the Crunchbase database, showed a significant lag of Russia from developed countries (0.52% of all startups in this area).

Keywords

big data; big data technologies; big data industry; monitoring

References

1. OECD (2011). OECD Guide to Measuring the Information Society 2011, OECD Publishing, Paris, <https://doi.org/10.1787/9789264113541-en>
2. Order of the Ministry of Communications and Mass Media of Russia No. 515 of December 7, 2015 "On Approval of the Collective Classification Groups "Information and Communication Technologies Sector" and "Content and Mass Media Sector" URL: <http://docs.cntd.ru/document/420327966> (accessed: 01.06.2021).
3. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture. URL: <https://www.iso.org/standard/71277.html> (accessed: 01.06.2021).
4. The European Data Market Study: Final Report. By IDC and Open Evidence. 2017. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirects=0&d=1 (accessed: 01.06.2021).
5. IDC's Worldwide Big Data and Analytics Spending Guide Taxonomy, IDC
6. Edinyj reestr rossijskih programm dlja jelektronnyh vychislitel'nyh mashin i baz dannyh. URL: <https://reestr.digital.gov.ru/> (accessed: 01.06.2021).
7. OECD (2013). SIZE OF THE ICT SECTOR. URL: <https://www.oecd-ilibrary.org/docserver/factbook-2011-72-en.pdf?expires=1632472701&id=id&acname=guest&checksum=8C61EC5E182618032918749217A03978> (accessed: 01.06.2021).

8. OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264311992-en>.
9. OECD Key ICT Indicators. URL: <https://www.oecd.org/sti/ieconomy/oecdkeyictindicators.htm>
10. Core list of indicators. March 2016 version. Partnership on Measuring ICT for Development. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf (accessed: 01.06.2021).
11. ICT sector (isoc_se). Reference Metadata in Euro SDMX Metadata Structure (ESMS). Eurostat, the statistical office of the European Union. URL: https://ec.europa.eu/eurostat/cache/metadata/en/isoc_se_esms.htm (accessed: 01.06.2021).
12. Eurostat. Browse statistics by theme. Digital economy and society. Database. ICT sector. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (accessed: 01.06.2021).
13. Indikatory cifrovoj jekonomiki: 2020: statisticheskij sbornik / G. I. Abdrahmanova, K. O. Vishnevskij, L. M. Gohberg i dr.; Nac. issled. un-t «Vysshaja shkola jekonomiki». – M.: NIU VShJe, 2020. URL: <https://issek.hse.ru/mirror/pubs/share/387609461.PDF> (accessed: 01.06.2021).
14. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European. - Luxembourg: Office for Official Publications of the European Communities, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (accessed: 01.06.2021).
15. T.V. Ershova, Y.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // Informatsionnoye obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
16. Rojas-Torres D., Kshetri N. Big data solutions for micro-, small-, and medium-sized enterprises in developing countries // IT Professional. 2019. Vol. 21. №. 5. P. 67-70. <https://doi.org/10.1109/MITP.2019.2932236>
17. Liang F. et al. A survey on big data market: Pricing, trading and protection // Ieee Access. 2018. Vol. 6. P. 15132-15154. <https://doi.org/10.1109/ACCESS.2018.2806881>
18. Ahmad M., Rashid T., Mishra D. K. Impact of wireless technology on future of big-data industry // 2014 Conference on IT in Business, Industry and Government (CSIBIG). IEEE, 2014. P. 1-4, <https://doi.org/10.1109/CSIBIG.2014.7056931>
19. Li C., Pan K. Research on financing efficiency of big data industry based on three stage DEA-Taking Guizhou Province as an example // Advances in Social Science, Education and Humanities Research (ASSEHR), 2018. Vol. 130. P. 439-442. <https://doi.org/10.2991/ifmeita-17.2018.74>
20. Zhang N., Zhang H., Zhang M. The Harmonious Development of Big Data Industry and Ecological Civilization Construction in Guizhou // 2018 6th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM 2018). Atlantis Press, 2019. P. 192-198. <https://doi.org/10.2991/ieesasm-18.2019.36>
21. Liu Y. et al. An overview of big data industry in China // China Communications. 2014. Vol. 11. № 12. P. 1-10. <https://doi.org/10.1109/CC.2014.7019834>
22. Kwon T. H., Kwak J. H., Kim K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea // Technological Forecasting and Social Change. – 2015. – T. 96. – S. 144-152. <https://doi.org/10.1016/j.techfore.2015.03.017>
23. Caputo F. et al. Over the mask of innovation management in the world of Big Data // Journal of Business Research. 2020. Vol. 119. P. 330-338. <https://doi.org/10.1016/j.jbusres.2019.03.040>
24. Tang C. *The data industry: The business and economics of information and big data*. Hoboken: John Wiley & Sons, 2016, 195 p.
25. Weng W. H., Lin W. T. Development Trends and Strategy Planning in Big Data Industry // Contemporary Management Research. 2014. Vol. 10. № 3. P. 203-214. <https://doi.org/10.7903/cmr.12288>
26. Crunchbase. 2021. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirets=0&d=1 (accessed: 01.06.2021).

27. GOST R 58048-2017 Transfer tehnologij. Metodicheskie ukazaniya po ocenke urovnja zrelosti tehnologij; GOST R 57194.1-2016 Transfer tehnologij. Obshhie polozheniya. URL: <https://docs.cntd.ru/document/1200158331> (accessed: 01.06.2021).
28. Dalle J. M., Besten M., Menon C. Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers, No. 2017/08. Paris: OECD Publishing, 2017. 29 p. <https://doi.org/10.1787/6c418d60-en>
29. Kézai P. K., Fischer S., Lados M. Smart economy and startup enterprises in the Visegrád Countries – A comparative analysis based on the crunchbase database // Smart Cities. 2020. Vol. 3. №. 4. P. 1477-1494. <https://doi.org/10.3390/smartcities3040070>
30. Akhmetova J. A. The Impact of Qualitative Determinants on Startup Investments in Russia // Nauka, tehnologii i biznes: sbornik materialov II mezhvuzovskoj zaochnoj konferencii aspirantov, soiskatelej i molodyh uchenyh, Moskva, 28–29 aprelja 2020 goda. Moskva: Moskovskij gosudarstvennyj tehničeskij universitet imeni N. Je. Baumana (nacional'nyj issledovatel'skij universitet), 2020. P. 35-41.
31. Inshakov M. O., Orlova A. A. Innovacionnye startapy v Rossii: problemy sozdaniya i marketingovogo prodvizheniya // Vestnik Volgogradskogo gosudarstvennogo universiteta. Serija 3: Jekonomika. Jekologija. 2014. №. 1. P. 66-76

Технологии информационного общества**ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ
ДАНЫМИ В РОССИЙСКОМ БИЗНЕСЕ****Ершов Петр Сергеевич**

*Институт развития информационного общества, руководитель Дирекции региональных программ
Москва, Российская Федерация
peter.erшов@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель Совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, Лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Аннотация

Предложены концептуальная схема и набор показателей для мониторинга и оценки уровня использования технологий работы с большими данными в бизнесе, учитывающие международные стандарты, разработки аналитических компаний и релевантные научные публикации. В концептуальную схему включены показатели, характеризующие уровень зрелости работы с большими данными, использование технологий работы с большими данными и барьеры для их внедрения в организациях предпринимательского сектора. Проведена пилотная апробация системы мониторинга, рассчитаны отдельные показатели на основе доступных статистических данных за 2020 год.

Ключевые слова

большие данные; аналитика больших данных; предпринимательский сектор; бизнес; использование больших данных; уровень зрелости работы с большими данными; барьеры на пути использования больших данных; BD4DE; Big Data for Digital Economy

Введение

Сегодня становится общепринятой точка зрения, что данные в цифровой форме формируют основу цифровой экономики, а технологии хранения и анализа больших данных находятся в ряду основных технологических драйверов современного этапа цифровой трансформации различных сфер деятельности [1-5].

Характерно в этой связи определение цифровой экономики, данное в Стратегии развития информационного общества в Российской Федерации до 2030 года: «Цифровая экономика – хозяйственная деятельность, в которой ключевым фактором производства являются данные в цифровом виде, обработка больших объемов и использование результатов анализа которых по сравнению с традиционными формами хозяйствования позволяют существенно повысить эффективность различных видов производства, технологий, оборудования, хранения, продажи, доставки товаров и услуг» [6].

© Ершов П.С., Хохлов Ю.Е., Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_300

Предпринимательский сектор традиционно является пионером в использовании новых технологий, и технологии работы с большими данными не являются исключением. После получения подтвержденных положительных эффектов от использования технологий работы с большими данными в отдельных отраслях соответствующий опыт перенимают и другие сферы деятельности. Как правило, наиболее развитыми в использовании цифровых технологий являются финансовый сектор и ИКТ-сектор, их опыт в итоге используется уже в менее технологичных отраслях и в социальной сфере.

Несмотря на широко распространенную высокую оценку значимости технологий работы с большими данными для развития цифровой экономики и включение задач по развитию и использованию сквозных (цифровых) технологий в стратегические и программные документы, в России сохраняется существенный дефицит сведений о состоянии дел с производством, использованием и воздействием технологий работы с большими данными на экономику страны. Включение в 2020 г. в основную форму №3-информ федерального статистического наблюдения за использованием цифровых технологий в организациях [7] нескольких вопросов, касающихся больших данных, только частично решает эту проблему, – остаются недоступными несколько важных показателей. Для полноценной оценки текущего уровня использования технологий работы с большими данными российским бизнесом необходимо учитывать уровень готовности бизнеса к использованию таких технологий, уровень проникновения (использования) технологий работы с большими данными в различные бизнес-процессы, а также оценивать существующие барьеры.

В статье представлены компоненты концептуальной схемы и методология мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации BD4DE (Big Data for Digital Economy) [8], относящиеся к оценке использования технологий работы с большими данными в российском бизнесе. Предложенная методология опирается на подходы, разработанные в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации» Национальной технологической инициативы.

1 Обзор международных стандартов и основных публикаций по мониторингу использования технологий работы с большими данными в бизнесе

Основной разработчик стандартов статистического наблюдения за использованием цифровых технологий – Организация экономического сотрудничества и развития (ОЭСР) – впервые включила вопросы об использовании аналитики больших данных в модельное обследование бизнеса в 2015 г. [9]. Большим данным там посвящен специальный раздел, содержащий рекомендации по сбору следующих сведений.

1. Использование аналитики [больших] данных – предприятия, проводящие анализ больших данных собственными силами или заказывающие анализ сторонним организациям.
2. Расходы на аналитику [больших] данных – доля от общих переменных затрат предприятия, включая расходы на персонал.
3. Мотивация для использования аналитики [больших] данных – (а) идентификация потенциальных клиентов; (б) увеличение расходов клиентов (за счет таргетирования предложений и скидок и т.д.); (в) адаптация продуктов к потребностям клиентов; (г) повышение эффективности внутреннего производства и/или организации (например, выявление узких мест, лучших практик и т. д.).
4. Воздействие аналитики [больших] данных – (а) экономия средств; (б) рост продаж (в том числе за счет улучшения продуктов и более эффективного маркетинга); (в) улучшения в организации бизнеса. При этом оценка воздействия предприятиями может выполняться в шкале «да / нет» или на основе качественной шкалы (например, [воздействие] высокое / частичное / ограниченное / отсутствует).
5. Барьеры, ограничивающие использование аналитики [больших] данных или препятствующие ему – (а) бесполезность технологий работы с большими данными в бизнесе; (б) ограниченность ожидавшейся прибыли в сравнении с затратами; (в) отсутствие навыков у персонала; (г) правовые проблемы и риски.

Евростат оперативно отреагировал на модельное обследование ОЭСР и включил раздел об использовании аналитики больших данных в свои модельные анкеты для бизнеса 2016, 2018 и 2020 гг. [10-12].

Показатели в обследованиях Евростата отличались от разработанных ОЭСР: из пяти показателей в анкете 2016 г. был включен только один об использовании аналитики данных (проведение анализа больших данных собственными силами или сторонней организацией), но добавлен новый показатель об источниках больших данных с возможностью выбора из следующих вариантов: (а) собственные данные от «умных» приборов и сенсоров; (б) геолокационные данные от мобильных устройств; (в) данные, генерируемые социальными медиа (социальные сети, блоги, мультимедийные онлайн-платформы и т.д.); (г) иные источники [9].

В 2019 г. модельная анкета Евростата (для обследования предприятий в 2020 г.) была доработана, и в нее вошли следующие показатели использования аналитики больших данных на предприятии [12].

1. Проведение анализа данных из различных источников (вопрос из обследования 2016 г.).
2. Используемые методы анализа больших данных с двумя вариантами ответов: (а) машинное обучение, в том числе глубокое обучение (deep learning); (б) обработка естественного языка, генерация естественного языка или распознавание речи.
3. Наличие у предприятия другой организации или предприятия, проводивших для него анализ больших данных.
4. Рассмотрение предприятием возможности использования аналитики больших данных (для тех, которые пока не используют).
5. Барьеры, ограничивающие или препятствующие использованию аналитики [больших] данных на предприятии (модифицированный вопрос ОЭСР для тех, кто не использует аналитику больших данных) – (а) затраты представляются слишком высокими по сравнению с выгодами; (б) недостаток человеческих ресурсов, знаний, навыков, например, требуемых специалистов недостаточно на предприятии, или их трудно нанять; (в) недостаточно источников больших данных как внутри, так и за пределами предприятия, которые были бы необходимы для анализа больших данных; (г) недостаточная ИКТ-инфраструктура – отсутствие соответствующего программного или аппаратного обеспечения для выполнения необходимых обработки и анализа; (д) трудности, связанные с соблюдением законов о защите персональных данных; (е) такой анализ не является приоритетом для предприятия; (ж) низкое качество источников больших данных; (з) анализ больших данных бесполезен для предприятия; (и) другие причины.
6. Продажа предприятием больших данных (или доступа к ним) в отчетном году.
7. Покупка предприятием больших данных (или доступа к ним) в отчетном году.

В 2020 г. с учетом модельных анкет ОЭСР и Евростата в форму федерального статистического наблюдения №3-Информ, по которой собираются сведения об использовании информационных и коммуникационных технологий, был включен ряд вопросов об использовании технологий работы с большими данными [7]. В форму вошли вопросы об использовании «технологий сбора, обработки и анализа больших данных», об источниках анализируемых данных и цели использования технологий их сбора, обработки и анализа, а также вопрос о том, какими силами проводится анализ больших данных (собственными силами или сотрудниками других организаций). Последний вопрос аналогичен вопросу из анкеты Евростата, второй соответствует ей лишь отчасти.

Основной проблемой международных стандартов и действующего инструментария статистического наблюдения с точки зрения проведения комплексного мониторинга развития и использования технологий работы с большими данными является то обстоятельство, что эти технологии рассматриваются как отдельный случай цифровых технологий и по ним не собираются, в частности, сведения для оценки различных аспектов зрелости работы с большими данными на предприятиях, включая комплекс вопросов относительно существующих условий для работы с большими данными (кадры, цифровая инфраструктура, организация работы с большими данными и др.). Хотя официального статистического инструментария на данный момент не существует, аналитическими компаниями и исследователями разработан ряд диагностических методологий определения уровня зрелости работы с большими данными в организациях, например [13-17]. Концептуальные схемы и показатели моделей зрелости работы с большими данными были учтены при разработке мониторинга использования технологий работы с большими данными, аналитический обзор этих моделей представлен в посвященной им статье этого номера [18].

Использование технологий работы с большими данными в различных отраслях бизнеса является предметом многочисленных научных исследований, опубликовано достаточно много

обзоров таких исследований. В качестве примера можно привести обзоры исследований использования технологий работы с большими данными в промышленности [19], сельском хозяйстве [20], строительстве [21], городской среде [22], транспорте [23], энергетической инфраструктуре [24], финансовом секторе [25]. Результаты этих исследований также учитывались при формировании системы показателей для мониторинга использования технологий работы с большими данными.

2 Описание концептуальной схемы мониторинга использования технологий работы с большими данными в бизнесе

Исходя из проведенного анализа международных стандартов статистического наблюдения, моделей зрелости и исследований использования технологий работы с большими данными в бизнесе в концептуальную схему мониторинга предлагается включить следующие области оценки (рисунок 1):

- уровень зрелости работы бизнеса с большими данными;
- использование технологий работы с большими данными в бизнесе;
- барьеры для использования технологий работы с большими данными в бизнесе.

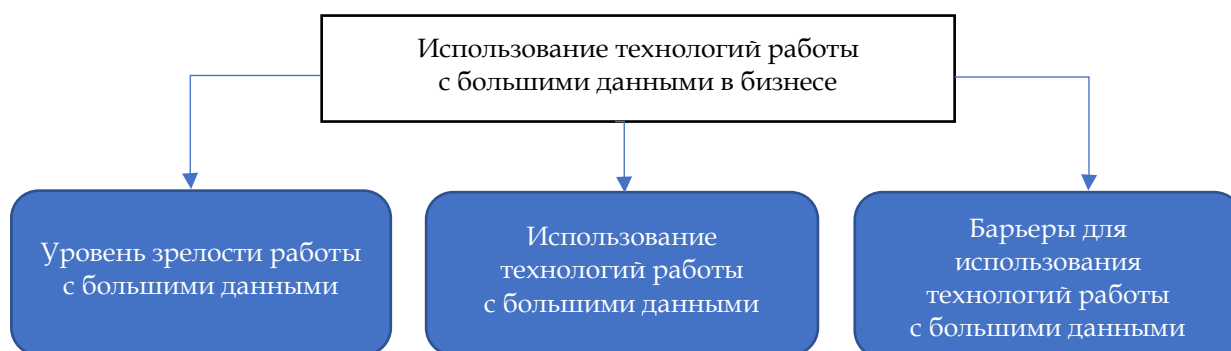


Рис. 1. Концептуальная схема использования технологий работы с большими данными в российском бизнесе

2.1 Уровень зрелости работы с большими данными в бизнесе

В основу набора показателей для мониторинга уровня зрелости работы с большими данными в бизнесе положена разработанная с участием авторов настоящей статьи модель зрелости BD4DE-MM работы с большими данными в организациях [18]. При этом в расчете показателей учитываются предприятия, достигшие продвинутого уровня зрелости, т.к. на более низких уровнях зрелости работа с большими данными практически не ведется.

Предлагается следующий набор показателей, характеризующих уровень зрелости работы с большими данными в бизнесе (показатели рассчитываются для бизнеса в целом, а также в разрезе отраслей и размеров предприятий).

- (УЗБ-01) Доля предприятий, имеющих продвинутый и выше уровень зрелости работы с большими данными

Показатель характеризует зрелость работы с большими данными в целом по бизнесу (или по отдельной отрасли) и может быть использован, в частности, для выявления отраслей-лидеров по работе с большими данными.

- (УЗБ-01-01) Доля предприятий, имеющих продвинутый и выше уровень зрелости стратегии и регулирования работы с большими данными

Показатель характеризует степень распространенности практики подготовки предприятиями проработанных документов стратегического планирования и регулирования использования технологий для работы с большими данными для достижения бизнес-целей.

- (УЗБ-01-02) Доля предприятий, имеющих продвинутый и выше уровень кадрового обеспечения работы с большими данными

Показатель характеризует степень обеспеченности предприятий предпринимательского сектора кадрами, подготовленными в области технологий работы с большими данными, а также наличие руководителей, заинтересованных во внедрении технологий работы с большими данными и имеющих соответствующие компетенции.

- (УЗБ-01-03) Доля предприятий, имеющих продвинутый и выше уровень обеспеченности качественными большими данными

Показатель характеризует степень использования больших данных в различных бизнес-процессах, а также качество данных, их интеграцию и совместное использование.

- (УЗБ-01-04) Доля предприятий, имеющих продвинутый и выше уровень использования аналитики и инструментов для работы с большими данными

Показатель характеризует использование специализированных инструментов для работы с большими данными, а также регулярность использования аналитики в различных бизнес-процессах.

- (УЗБ-01-05) Доля предприятий, имеющих продвинутый и выше уровень зрелости цифровой инфраструктуры и безопасности для работы с большими данными

Показатель характеризует наличие и возможности масштабирования цифровой инфраструктуры, необходимой для работы с большими данными, а также уровень доверия и безопасности работы с большими данными.

- (УЗБ-01-06) Доля предприятий, имеющих продвинутый и выше уровень зрелости организационных механизмов работы с большими данными

Показатель характеризует наличие у предприятий необходимых организационных единиц (ответственного руководителя, профильного подразделения, центра компетенции) для работы с большими данными.

2.2 Использование технологий работы с большими данными в бизнесе

Хотя вопросы использования технологий работы с большими данными затрагиваются в модели зрелости (и основанных на ней показателях), важно проводить мониторинг по всему спектру ключевых показателей использования технологий работы с большим данными в бизнесе. Представленный далее набор показателей, в ряду прочего, формировался и с учетом разработанных международных стандартов и доступности данных, что позволяет проводить международные сопоставления по отдельным показателям.

- (ИБДБ-01) Доля предприятий, использующих технологии для работы с большими данными

Показатель характеризует масштабы использования технологий работы с большими данными.

- (ИБДБ-02) Доля предприятий, выполняющих анализ больших данных собственными силами или силами внешних организаций

Показатель характеризует особенности работы с большими данными с точки зрения наличия для этого необходимых инструментов и компетенций внутри организации.

- (ИБДБ-03) Доля предприятий, продающих или предоставляющих доступ к большим данным

Показатель характеризует количество поставщиков на рынке больших данных.

- (ИБДБ-04) Доля предприятий, покупающих или получающих доступ к большим данным

Показатель характеризует заинтересованность предприятий в обогащении имеющихся у них данных большими данными сторонних организаций.

- (ИБДБ-05) Доля затрат предприятий на работу с большими данными в общих затратах на использование цифровых технологий

Показатель характеризует финансовую обеспеченность использования технологий для работы с большими данными.

- (ИБДБ-06) Доля предприятий, применяющих аналитику больших данных по целям использования

Данный показатель характеризует направления использования аналитики больших данных с целью повышения эффективности бизнес-процессов, таких как:

- маркетинг и увеличение продаж;
- производственные процессы;

- административные процессы;
- управление предприятием;
- логистика;
- обеспечение информационной безопасности;
- управление кадрами.

2.3 Барьеры для работы с большими данными

Для мониторинга и оценки использования технологий работы с большими данными необходимо отслеживать не только текущий уровень этого использования. Важным для мониторинга как инструмента управления является наличие показателей, характеризующих причины, по которым предприятия не используют технологии работы с большими данными.

Представленные в этом разделе показатели основаны на вопросах, которые задаются предприятиям, не использующим технологии работы с большими данными, с просьбой указать препятствия, мешающие внедрению этих технологий. Такие вопросы отсутствуют в форме федерального статистического наблюдения №3-информ, их измерение возможно на основе представительного опроса предприятий. Показатели также разрабатывались с учетом возможности международного сопоставления.

- (ББДБ-01) Доля предприятий, считающих, что работа с большими данными не является приоритетом для их деятельности

Понимание выгод от использования технологий работы с большими данными является важным фактором их использования. Показатель характеризует, насколько высоко предприятия оценивают значимость и возможную пользу от внедрения технологий работы с большими данными, определяя приоритеты своей деятельности.

- (ББДБ-02) Доля предприятий, считающих, что затраты на работу с большими данными слишком высоки по сравнению с выгодами

Любая технология на этапе своего внедрения требует инвестиций, и если крупные компании могут позволить себе эксперименты с новыми технологиями (которые на начальных этапах обычно дороги), то небольшие организации действуют, как правило, с отставанием, особенно если эффекты для организации не очевидны. Задача данного показателя – продемонстрировать, какая доля компаний не готова тратить деньги на внедрение технологий работы с большими данными, считая, что инвестиции себя не окупят.

- (ББДБ-03) Доля предприятий, считающих, что работа с большими данными бесполезна для организации

Показатель характеризует наличие компаний, которые не видят выгод от использования технологий работы с большими данными в своей деятельности.

- (ББДБ-04) Доля предприятий, считающих, что у них недостаточно человеческих ресурсов, знаний, навыков для работы с большими данными

Наличие подготовленных кадров является важным фактором при внедрении новых технологий, а их отсутствии несомненно является барьером. Данный показатель характеризует недостаток человеческих ресурсов, знаний и навыков, необходимых для работы с большими данными.

- (ББДБ-05) Доля предприятий, считающих, что у них недостаточно развита цифровая инфраструктура для работы с большими данными

Технологии для работы с большими данными требовательны к цифровой инфраструктуре, которую могут себе позволить далеко не все предприятия. Данный показатель характеризует число компаний, чья текущая инфраструктура не подходит для внедрения технологий для работы с большими данными и требует существенного изменения.

- (ББДБ-06) Доля предприятий, считающих, что источников больших данных для решения текущих задач организации недостаточно

Для того чтобы извлечь дополнительную ценность из имеющихся данных, исходные массивы данных должны, как правило, быть обогащены данными из других источников, что не всегда возможно. Данный показатель характеризует распространенность данного препятствия использования технологий работы с большими данными.

- (ББДБ-07) Доля предприятий, считающих, что качество имеющихся больших данных как внутри, так и за пределами организации недостаточно для решения нужных задач

Наличие большого количества источников данных является необходимым, но не достаточным условием для решения бизнес-задач; важнейшим параметром больших данных является их качество. Каждая организация самостоятельно определяет этот критерий, но, как правило, он состоит из таких свойств, как актуальность, полнота, целостность, объективность, релевантность и т.д. В общем случае качество данных, в том числе больших – это степень их пригодности к обработке и анализу, а также соответствие обязательным и специальным требованиям. Данный показатель характеризует число компаний, для которых качество доступных больших данных недостаточно для решения своих бизнес-задач с использованием технологий работы с ними.

- (ББДБ-08) Доля предприятий, считающих, что законы о персональных данных и их конфиденциальности мешают работе с большими данными

Закон о персональных данных запрещает обработку персональных данных субъекта без его явного согласия и указания целей обработки, что ограничивает возможности организаций по сбору и обработке пользовательских данных. Данный показатель характеризует законодательство о персональных данных как барьер при использовании технологий для работы с большими данными.

- (ББДБ-09) Доля предприятий, имеющих опасения относительно правовых последствий работы с большими данными

Современные методы обработки больших данных успешно решают задачу привязки персонифицированных данных к конкретному физическому лицу путем объединения данных из разных источников, каждый из владельцев которых формально не нарушает законодательство и использует процедуры деперсонификации. Это приводит к появлению так называемой серой зоны, работать в которой согласны не все предприятия. Данный показатель характеризует наличие компаний, опасющихся правовых последствий от работы с большими данными.

- (ББДБ-10) Доля организаций, считающих использование аналитики больших данных не этичным

Современные методы работы с большими данными способны помогать решать бизнесу не только задачи, связанные с бизнесом, но и вторгаться в личную жизнь человека, что ставит вопрос об этичности использовании технологий работы с большими данными и может выступать препятствием для их использования.

- (ББДБ-11) Доля организаций, указавших другие причины

Т.к. изучение барьеров, препятствующих внедрению в деятельность организаций технологий работы с большими данными, носит сложный и часто междисциплинарный характер, для своевременного обновления статистического инструментария важно контролировать, насколько предложенный спектр барьеров покрывает все причины, по которым предприятия не используют эти технологии.

3 Методология мониторинга

Для измерения показателей, характеризующих уровень зрелости работы с большими данными на предприятиях, использование технологий работы с большими данными, а также существующие барьеры, необходимо обеспечить сбор данных, для чего разработан соответствующей инструментарий в виде анкеты, содержащей модельные вопросы по каждому показателю и необходимые инструкции по заполнению, включающие пояснения и определения. Сбор данных должен проводиться в рамках выборочных представительных опросов организаций предпринимательского сектора разного размера (малые, средние, большие) из разных отраслей. Для отбора организаций предпринимательского сектора должна использоваться случайная стратифицированная выборка. Подробное описание методологии сбора данных и расчета значений показателей для отдельной сферы деятельности приведено в статье [8].

Разработанная для опроса анкета может быть легко реализована в онлайн-формате, а опрос может проводиться с использованием электронных средств коммуникации, что существенно ускорит сбор данных.

Респонденты должны подбираться из числа лиц, принимающих решения в организации по различным направлениям: стратегическое планирование развития организации, кадровое

обеспечение, использование ИКТ, информационная безопасность и доверие. Разработанная анкета содержит рекомендации по выбору респондента по каждому разделу анкеты.

Для отдельных показателей данные берутся из результатов федерального статистического наблюдения, проведенного Росстатом по новой форме №3-информ, содержащей соответствующие вопросы.

Для расчета значения всех показателей были разработаны паспорта, содержащие алгоритмы расчета значений, разрезы и весовые коэффициенты.

4 Результаты

Учитывая особенности разработанного состава показателей, получение полного набора данных по всем трем направлениям предлагаемой концептуальной схемы возможно после проведения опроса предприятий. Вместе с тем в настоящее время стали доступны результаты федерального статистического наблюдения по форме №3-информ за 2020 г. [26], на основе которых проведены расчеты отдельных показателей в рамках пилотной реализации методологии мониторинга.

По показателю «Доля предприятий, использующих технологии для работы с большими данными» его значения в разрезе отраслей представлены на рисунке 2.



Рис. 2. Доля предприятий различных отраслей бизнеса, использующих технологии сбора, обработки и анализа больших данных, 2020, %

Наиболее активно используют технологии работы с большими данными предприятия финансового сектора (включая страхование) – 44,4%. И это понятно: инструменты работы с большими данными им нужны для хранения и обработки больших массивов данных, оценки рисков, обеспечения безопасности и т.д. В ИКТ-секторе таких предприятий около трети, его лидерство связано с наличием необходимых компетенций и доступностью технологий. Аутсайдерами в использовании технологий работы с большими данными являются предприятия отраслей сельского хозяйства и строительства (17,2% и 16,3% соответственно).

Значение показателя «Доля предприятий, выполняющих анализ больших данных собственными силами или силами внешних организаций» для различных отраслей представлены на рисунке 3.

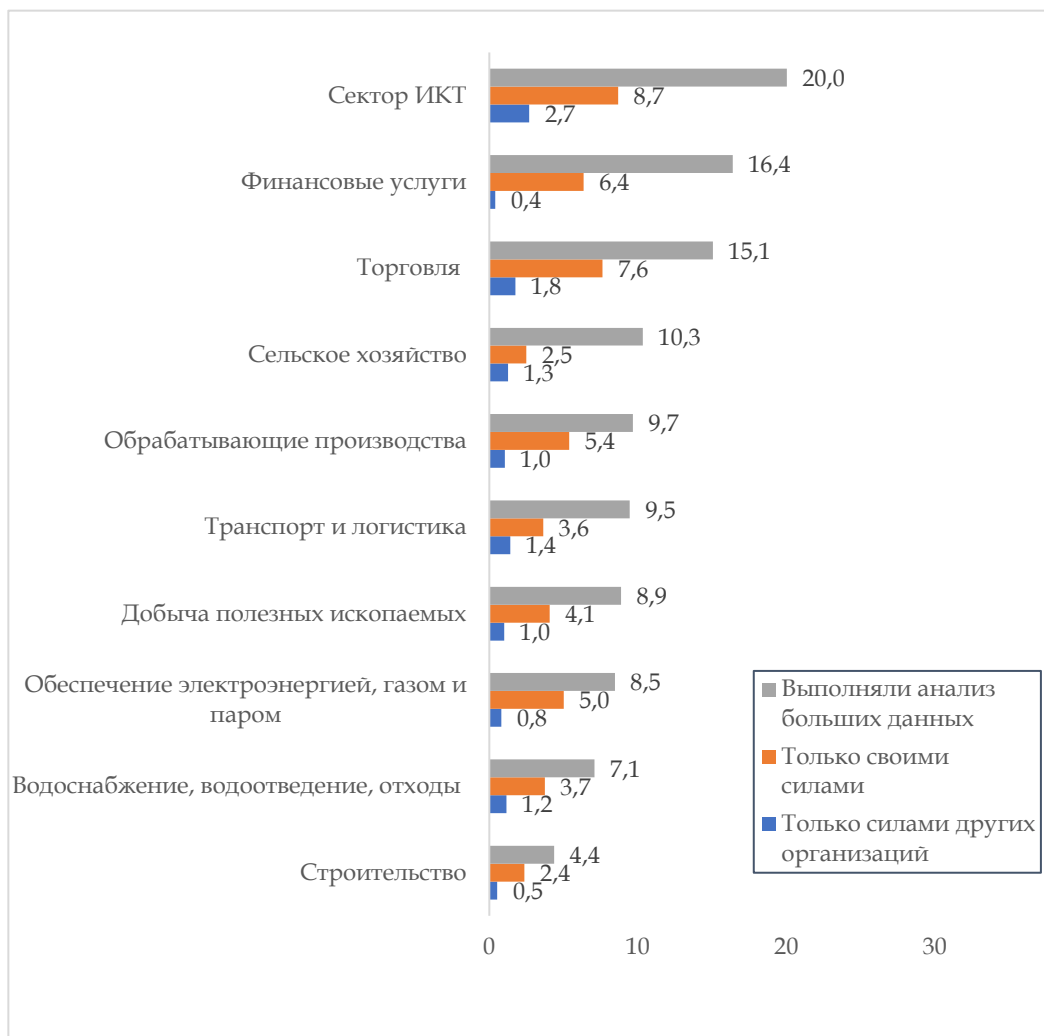


Рис. 3. Доля предприятий различных отраслей, выполнявших анализ больших данных, 2020, %

Лидером по проведению анализа больших данных только собственными силами (8,7%) является ИКТ-сектор, который также наиболее активно использует аналитику больших данных (20%). Можно предположить, что это лидерство в значительной степени обеспечено наличием компетенций и инфраструктуры для работы с большими данными. Высокий уровень использования аналитики больших данных у финансового сектора и торговли. Важным использованием аналитики больших данных является и для сельского хозяйства (предсказание урожая, прогноз цен на рынках, обработка результатов мониторинга и т.д.). Аутсайдером, как и по первому показателю, является строительство: только 4,4% проводят анализ больших данных и 0,5% – исключительно своими силами.

На рисунке 4 приведены данные для ЕС по близкому показателю «Доля предприятий, проводящих анализ больших данных», рассчитанному в целом по бизнесу в разрезе стран ЕС. Среднее значение для ЕС в 2020 году составляет 13%, страны-лидеры достигают по этому показателю значений в 20–30% (Мальта – 31%, Дания, Нидерланды и Великобритания по 27%, Бельгия и Ирландия по 23%), аутсайдерами рейтинга в основном являются страны Восточной Европы. При сравнении и интерпретации этих данных необходимо учитывать, что в ЕС данные получают на основе представительного опроса предприятий, включая малые с числом занятых 10 человек и более (малые – это основной массив предприятий), тогда как в России по форме №3-информ проводится сплошное обследование, которое не охватывает предприятия, имеющие статус малых, что приводит к смещению совокупности обследованных предприятий в сторону средних и крупных, более широко использующих технологии и аналитику больших данных.

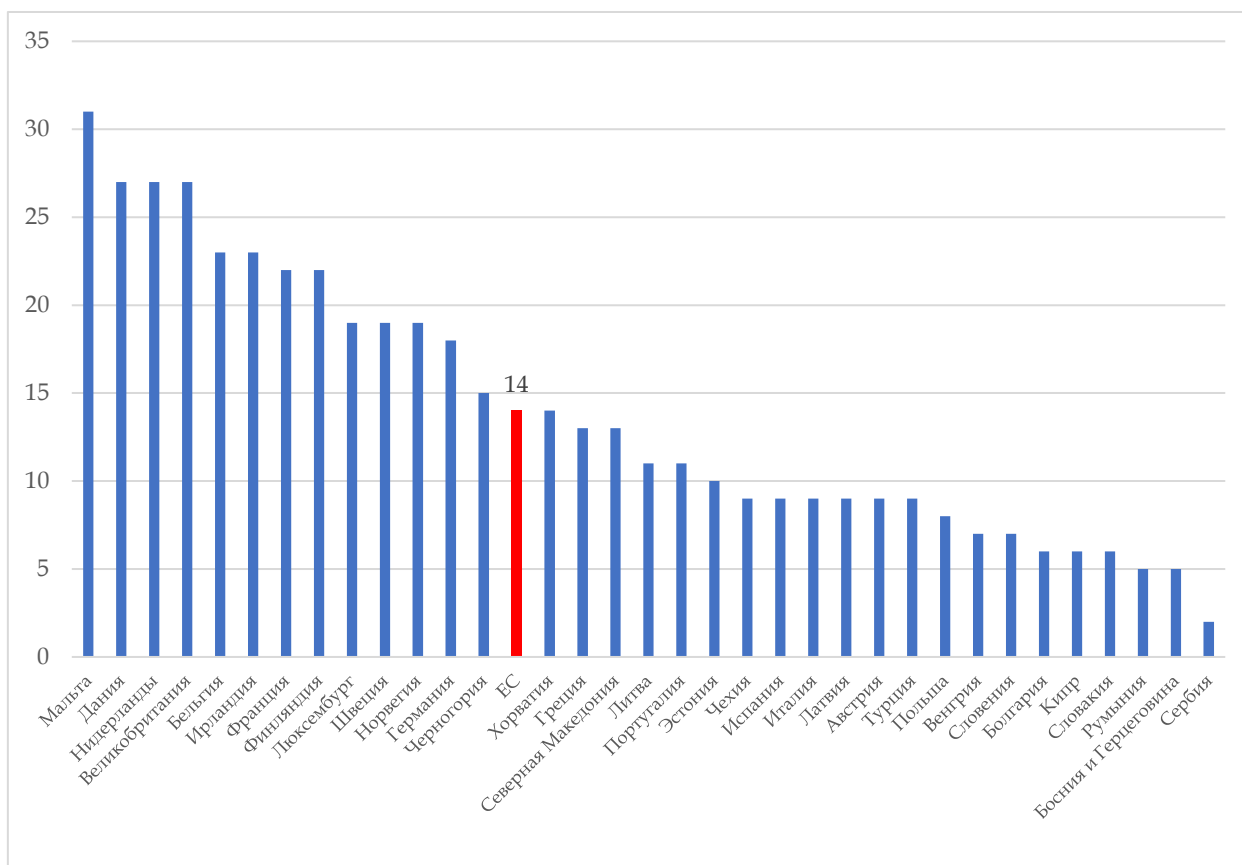


Рис. 4. Доля предприятий, проводящих анализ больших данных в странах ЕС, 2020, %
Источник: [27]

Заключение

Представленная система показателей для мониторинга использования технологий работы с большими данными в российском бизнесе состоит из 23 показателей. Основную сложность для измерения представляют показатели для мониторинга уровня зрелости работы с большими данными, т.к. необходимые для их расчета данные могут быть получены только на основе опроса предприятий и требуют привлечения к заполнению опросных форм представителей различных служб. Показатели использования технологий работы с большими данными частично могут быть получены из результатов статистического наблюдения по форме №3-информ и имеют (ограниченную) возможность сопоставления с данными по странам ЕС. Показатели для мониторинга барьеров использования больших данных тоже можно получить на основе опроса предприятий, и здесь имеется широкая возможность для сопоставления с данными ЕС.

В качестве дальнейшего развития настоящего исследования необходимо провести комплекс мероприятий по получению соответствующих данных на основе разработанного инструментария с достаточной выборкой по разным отраслям с целью апробации разработанной системы показателей для мониторинга использования технологий работы с большими данными в российском бизнесе. Кроме того, необходимо вести работу по регулярному отслеживанию изменений в системе статистических наблюдений организаций разного уровня (МСЭ, ОЭСР, Евростат) с целью отбора релевантных показателей и актуализации системы мониторинга.

Полученные в ходе пилота данные говорят о достаточно широком использовании технологий работы с большими данными и аналитики больших данных в бизнесе при существенной дифференциации отраслей по этим показателям (значения показателей лидеров и аутсайдеров различаются в разы). Большая часть предприятий, использующих аналитику больших данных, проводит анализ самостоятельно (полностью или частично, передавая часть работы сторонним

организациям), полностью на аутсорсинг отдает анализ больших данных малое число предприятий.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. Всемирный банк. 2021 год. Доклад о мировом развитии 2021 «Данные для лучшей жизни». Всемирный банк, Вашингтон, округ Колумбия.
2. Raquet Eric, Viktor Herna, Guo Hongyu. Data Mining in Finance: Current Advances and Future Challenges. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000369157100008> (дата обращения: 01.06.2021)
3. Rossell David. BIG DATA AND STATISTICS A STATISTICIAN'S PERSPECTIVE. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000366726700021> (дата обращения: 01.06.2021)
4. Ahmad Javaria, Yousef Amr, Buckles Bill. Movie Success Prediction Using Data Mining. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000425956500275> (дата обращения: 01.06.2021)
5. Mansingh Gunjan, Rao Lila, McNaughton Maurice. Discovering Knowledge Nuggets in Financial Data: The Case of a Financial Services Institution. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000425541200053> (дата обращения: 01.06.2021)
6. О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы. Указ Президента Российской Федерации от 9 мая 2017 г. № 203. URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102431687> (дата обращения: 01.06.2021)
7. Приказ Росстата N 424 от 30.07.2020 "Об утверждении форм федерального статистического наблюдения для организации федерального статистического наблюдения за деятельностью в сфере образования, науки, инноваций и информационных технологий". Приложение. URL: https://gks.ru/bgd/free/b16_27/IssWWW.exe/Stg/d05/pril_424_1.doc (дата обращения: 01.06.2021)
8. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
9. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (дата обращения: 01.06.2021)
10. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2016. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/a39ae859-8a16-4306-8020-ae06d3df3c91/Questionnaire%20ENT%202016.pdf> (дата обращения: 01.06.2021)
11. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2018. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/1fbef4a1-4c31-4b6a-afe8-19ee6d7e3b0f/ICT-Entr%202018%20-%20Model%20Questionnaire%20V%201.2.pdf> (дата обращения: 01.06.2021)

12. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2020. General outline of the survey. URL: https://circabc.europa.eu/sd/a/47b2dcfa-2eb9-4cc4-9e98-b93a85406d67/MQ_2020 ICT_ENT.pdf (дата обращения: 01.06.2021)
13. IBM Big Data & Analytics Maturity Model. URL: https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en (дата обращения 19.11.2020)
14. TDWI. A Guide to Achieving Big Data Analytics Maturity (v 1.1.2018). URL: <https://tdwi.org/whitepapers/2018/01/adv-all-ms-a-guide-to-achieving-big-data-analytics-maturity.aspx> (дата обращения 19.11.2020)
15. Hortonworks Big Data Maturity Model. A Hortonworks White Paper. URL: <http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf> (дата обращения 19.11.2020)
16. IDC MaturityScape Benchmark: Big Data and Analytics in the United States. URL: <https://tdwi.org/whitepapers/2016/06/cloudera-idc-big-data-and-analytics-in-the-us.aspx> (дата обращения 19.11.2020)
17. TDWI Advanced Analytics Maturity Model Assessment. URL: <https://tdwi.org/whitepapers/2018/04/adv-all-tdwi-advanced-analytics-maturity-model-guide.aspx> (дата обращения 19.11.2020).
18. А.В. Катин, П.С. Ершов, Ю.Е. Хохлов, С.Б. Шапошник Модель BD4DE-MM зрелости работы с большими данными в организации // Информационное общество. 2021, № 4–5. С. 259–277. https://doi.org/10.52605/16059921_2021_04_259
19. Yang Jing, Li Shaobo, Wang Zheng, Dong Hao, Wang Jun, Tang Shihao. Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges. <https://doi.org/10.3390/ma13245755>
20. Kamilaris Andreas, Kartakoullis Andreas, Prenafeta-Boldu Francesc. A review on the practice of big data analysis in agriculture. <https://doi.org/10.1016/j.compag.2017.09.037>
21. Huang Yao, Shi Qian, Zuo Jian, Pena-Mora Feniosky, Chen Jindao. Research Status and Challenges of Data-Driven Construction Project Management in the Big Data Context. <https://doi.org/10.1155/2021/6674980>
22. Niska Harri, Serkkola Ari. Data analytics approach to create waste generation profiles for waste management and collection. <https://doi.org/10.1016/j.wasman.2018.04.033>
23. Lyu Tao, Wang Peirong, Gao Yanan, Wang Yuanqing. Research on the big data of traditional taxi and online car-hailing: A systematic review. <https://doi.org/10.1016/j.jtte.2021.01.001>
24. Ahmad, Tanveer, Chen Huanxin, Wang Jianguy, Guo Yabin. Review of various modeling techniques for the detection of electricity theft in smart grid environment. <https://doi.org/10.1016/j.rser.2017.10.040>
25. Andriosopoulos Dimitris, Doumpos Michalis, Pardalos Panos M., Zopounidis Constantin. Computational approaches and data analytics in financial services: A literature review. <https://doi.org/10.1080/01605682.2019.1595193>
26. Федеральная служба государственной статистики. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании информационных и коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказании услуг в этих сферах». URL: <https://rosstat.gov.ru/folder/14478> (дата обращения 15.10.2021)
27. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (дата обращения 01.06.2021)

USE OF BIG DATA TECHNOLOGIES IN RUSSIAN BUSINESS

Ershov, Peter Sergeevich

*Institute of the Information Society, head of Directorate of regional programs
Moscow, Russian Federation
peter.ershov@iis.ru*

Hohlov, Yuri Evgenyevich

*Candidate of Physical and Mathematical Sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Shaposhnik, Sergei Borisovich

*Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for
regional development, senior researcher
Petrozavodsk, Russian Federation
sergei.shaposhnik@gmail.com*

Abstract

A framework and a set of indicators are proposed for monitoring and assessing the level of use of big data technologies in business, taking into account international standards, developments of analytical companies and relevant scientific publications in this area. The framework includes indicators characterizing the level of maturity of working with big data, the use of big data technologies and barriers to their implementation in companies. A pilot approbation of the monitoring system was carried out, individual indicators were calculated based on the available statistical data for 2020.

Keywords

big data; big data analytics; business; the use of big data by business; big data maturity level in business; barriers to the use of big data; BD4DE; Big Data for Digital Economy

References

1. Vsemirnyy bank. 2021 god. Doklad o mirovom razvitii 2021 «Dannyye dlya luchshey zhizni». Vsemirnyy bank, Vashington, okrug Kolumbiya.
2. Paquet Eric, Viktor Herna, Guo Hongyu. Data Mining in Finance: Current Advances and Future Challenges. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000369157100008> (accessed on 01.06.2021)
3. Rossell David. BIG DATA AND STATISTICS A STATISTICIAN'S PERSPECTIVE. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000366726700021> (accessed on 01.06.2021)
4. Ahmad Javaria, Yousef Amr, Buckles Bill. Movie Success Prediction Using Data Mining URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000425956500275> (accessed on 01.06.2021)
5. Mansingh Gunjan, Rao Lila, McNaughton Maurice. Discovering Knowledge Nuggets in Financial Data: The Case of a Financial Services Institution URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000425541200053> (accessed on 01.06.2021)
6. O Strategii razvitiya informatsionnogo obshchestva v Rossiyskoy Federatsii na 2017 - 2030 gody. Ukaz Prezidenta Rossiyskoy Federatsii ot 9 maya 2017 g. № 203. URL: <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102431687> (accessed on 01.06.2021)
7. Prikaz Rosstata N 424 ot 30.07.2020 "Ob utverzhdenii form federal'nogo statisticheskogo nablyudeniya dlya organizatsii federal'nogo statisticheskogo nablyudeniya za deyatel'nost'yu v sfere obrazovaniya, nauki, innovatsiy i informatsionnykh tekhnologiy". Prilozheniye. URL:

- https://gks.ru/bgd/free/b16_27/IssWWW.exe/Stg/d05/pril_424_1.doc (accessed on 01.06.2021)
8. T.V. Yershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // Informatsionnoye obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
 9. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. OECD 2015. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (accessed on 01.06.2021)
 10. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2016. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/a39ae859-8a16-4306-8020-ae06d3df3c91/Questionnaire%20ENT%202016.pdf> (accessed on 01.06.2021)
 11. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2018. General outline of the survey. URL: <https://circabc.europa.eu/sd/a/1fbef4a1-4c31-4b6a-afe8-19ee6d7e3b0f/ICT-Entr%202018%20-%20Model%20Questionnaire%20V%201.2.pdf> (accessed on 01.06.2021)
 12. COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES 2020. General outline of the survey. URL: https://circabc.europa.eu/sd/a/47b2dcfa-2eb9-4cc4-9e98-b93a85406d67/MQ_2020_ICT_ENT.pdf (accessed on 01.06.2021)
 13. IBM Big Data & Analytics Maturity Model. URL: https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en (accessed on 19.11.2020)
 14. TDWI. A Guide to Achieving Big Data Analytics Maturity (v 1.1.2018). URL: <https://tdwi.org/whitepapers/2018/01/adv-all-ms-a-guide-to-achieving-big-data-analytics-maturity.aspx> (accessed on 19.11.2020)
 15. Hortonworks Big Data Maturity Model. A Hortonworks White Paper. URL: <http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf> (accessed on 19.11.2020)
 16. IDC MaturityScape Benchmark: Big Data and Analytics in the United States. URL: <https://tdwi.org/whitepapers/2016/06/cloudera-idc-big-data-and-analytics-in-the-us.aspx> (accessed on 19.11.2020)
 17. TDWI Advanced Analytics Maturity Model Assessment. URL: <https://tdwi.org/whitepapers/2018/04/adv-all-tdwi-advanced-analytics-maturity-model-guide.aspx> (accessed on 19.11.2020).
 18. A.V. Katin, P.S. Yershov, YU.Ye. Khokhlov, S.B. Shaposhnik Model' BD4DE-MM zrelosti raboty s bol'shimi dannymi v organizatsii // Informatsionnoye obshchestvo. 2021, № 4–5. S. 259–277. https://doi.org/10.52605/16059921_2021_04_259
 19. Yang Jing, Li Shaobo, Wang Zheng, Dong Hao, Wang Jun, Tang Shihao. Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges. <https://doi.org/10.3390/ma13245755>
 20. Kamilaris Andreas, Kartakoullis Andreas, Prenafeta-Boldu Francesc. A review on the practice of big data analysis in agriculture. <https://doi.org/10.1016/j.compag.2017.09.037>
 21. Huang Yao, Shi Qian, Zuo Jian, Pena-Mora Feniosky, Chen Jindao. Research Status and Challenges of Data-Driven Construction Project Management in the Big Data Context. <https://doi.org/10.1155/2021/6674980>
 22. Niska Harri, Serkkola Ari. Data analytics approach to create waste generation profiles for waste management and collection. <https://doi.org/10.1016/j.wasman.2018.04.033>
 23. Lyu Tao, Wang Peirong, Gao Yanan, Wang Yuanqing. Research on the big data of traditional taxi and online car-hailing: A systematic review. <https://doi.org/10.1016/j.jtte.2021.01.001>
 24. Ahmad, Tanveer, Chen Huanxin, Wang Jiangyu, Guo Yabin. Review of various modeling techniques for the detection of electricity theft in smart grid environment. <https://doi.org/10.1016/j.rser.2017.10.040>
 25. Andriosopoulos Dimitris, Doumpos Michalis, Pardalos Panos M., Zopounidis Constantin. Computational approaches and data analytics in financial services: A literature review. <https://doi.org/10.1080/01605682.2019.1595193>
 26. Federal'naya sluzhba gosudarstvennoy statistiki. Itogi federal'nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol'zovanii informatsionnykh i kommunikatsionnykh tekhnologiy i proizvodstve vychislitel'noy tekhniki, programmnoy

- обеспечения и оказания услуг в этих сферах». URL: <https://rosstat.gov.ru/folder/14478> (accessed on 15.10.2021)
27. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (accessed on 01.06.2021)

Доверие и безопасность в информационном обществе

ДОВЕРИЕ И БЕЗОПАСНОСТЬ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ В РОССИИ

Катин Александр Владимирович

*Институт развития информационного общества, руководитель дирекции отраслевых программ РЭУ имени Г. В. Плеханова, старший преподаватель базовой кафедры цифровой экономики ИРИО Москва, Российская Федерация
alexander.katin@iis.ru*

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО Москва, Российская Федерация
yuri.hohlov@iis.ru*

Аннотация

Исследованы вопросы мониторинга и оценки уровня развития доверия и безопасности работы с большими данными в России. Разработана концептуальная схема и система показателей для мониторинга мер, предпринимаемых гражданами и организациями по обеспечению доверия и безопасности при работе с большими данными. Для оценки применимости разработанной концептуальной схемы проведена пилотная реализация.

Ключевые слова

большие данные, технологии работы с большими данными, информационная безопасность, доверие

Введение

Вопросы обеспечения информационной безопасности и доверия при использовании цифровых технологий в целом и технологий работы с большими данными в частности являются одними из важнейших факторов, влияющих на процессы цифровой трансформации экономики и общества. Получение ожидаемых социальных и экономических эффектов возможно только в том случае, когда технологии, решения и услуги, основанные на использовании больших данных, безопасны, а пользователи им доверяют.

В Российской Федерации этим вопросам традиционно уделяется повышенное внимание, что нашло свое отражение как в недавно утвержденной Стратегии национальной безопасности Российской Федерации [1], так и в действующей Доктрине информационной безопасности [2], представляющей систему официальных взглядов на обеспечение национальной безопасности Российской Федерации в информационной сфере. Также значимость вопросов обеспечения информационной безопасности отмечается в Стратегии развития информационного общества Российской Федерации до 2030 года [3].

Информационная безопасность Российской Федерации определяется как состояние защищенности личности, общества и государства от внутренних и внешних информационных угроз, при котором обеспечиваются реализация конституционных прав и свобод человека и гражданина, достойные качество и уровень жизни граждан, суверенитет, территориальная целостность и устойчивое социально-экономическое развитие Российской Федерации, оборона и безопасность государства [2]. В рамках настоящей статьи информационная безопасность будет рассматриваться только на уровне процессов, связанных с обеспечением безопасности и доверия

© Катин А.В., Хохлов Ю.Е., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2021_04_315

при использовании цифровых технологий в целом и технологий работы с большими данными в частности. Под доверием будет пониматься готовность пользователей (граждан и организаций) к полноценному применению цифровых технологий и решений с учетом всех возможных угроз и рисков.

Несмотря на наличие стратегических документов, высокий уровень вовлеченности лиц, принимающих решения со стороны органов власти, и выделение значительных финансовых ресурсов [4], в России отсутствует полноценная система мониторинга и оценки уровня доверия и безопасности работы с цифровыми технологиями в целом и с технологиями работы с большими данными в частности. Фактически публичная часть процесса мониторинга ограничивается опросами граждан и организаций в рамках федерального статистического наблюдения [5, 6], связанными лишь с отдельными аспектами информационной безопасности и доверия при использовании цифровых технологий, что не дает возможности проводить даже косвенную оценку уровня доверия и безопасности работы с большими данными.

Информационная безопасность существенно влияет на использование технологий работы с большими данными, поскольку, с одной стороны, может являться барьером (в силу своей сложности и инновационности), а с другой – фактором повышения уровня доверия пользователей (в условиях обеспечения гарантированно высокого уровня безопасности).

Целью данной статьи является разработка компонентов концептуальной схемы мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации (Big Data for Digital Economy, BD4DE) [7], относящихся к оценке уровня доверия и безопасности при работе с большими данными, на основе которой для отдельных показателей мониторинга получены оценки данной сферы в 2020 году.

1 Определение предметной области мониторинга и оценки

Бурное развитие и использование технологий работы с большими данными порождает множество вызовов для обеспечения информационной безопасности. Сверхбольшие объемы данных, их разнообразие и высокая скорость обработки порождают дополнительные риски, зачастую не возникавшие при использовании традиционных ИКТ.

Предметная область мониторинга в рамках настоящей статьи включает в себя два основных компонента: доверие и безопасность при работе с большими данными для граждан и то же самое для организаций.

Уровень доверия и безопасности граждан при работе с большими данными имеет значительный интерес для исследования, поскольку граждане, с одной стороны, являются пользователями технологий, решений и сервисов, основанных на работе с большими данными, а с другой – поставщиками данных. Низкий уровень безопасности данных граждан (прежде всего персональных) порождает множество рисков, связанных с финансовыми и репутационными потерями как для отдельных компаний-владельцев цифровых платформ и сервисов, так и для национальной экономики в целом, поскольку небезопасный оборот персональных данных приводит к повышению уровня преступности в цифровой среде. Слабый уровень доверия граждан к цифровым сервисам может приводить к отказу от их использования, что также является серьезным барьером для развития цифровой экономики.

Доверие и безопасность организаций при работе с большими данными также имеет важное значение, так как технологии, решения и сервисы, основанные на работе с большими данными, все глубже проникают в разнообразные деловые процессы, в то время как традиционные методы защиты информации, применяемые в организациях, не способны отвечать на все вызовы, порождаемые использованием больших данных в повседневной деятельности. Угрозы информационной безопасности, связанные с потерей или несанкционированным раскрытием данных организации, влекут за собой серьезные риски, поскольку, с одной стороны, они способны привести к финансовым и репутационным потерям, а с другой – свести к нулю ожидаемые эффекты от инвестиций в работу с большими данными в организации.

Развитие и использование технологий работы с большими данными в силу многих причин происходит неравномерно и имеет отраслевую специфику, поэтому представляется целесообразным проводить мониторинг и оценку уровня доверия и безопасности работы с большими данными в различных сферах деятельности. Это также важно для проведения межотраслевых сопоставлений. Данные сведения будут полезны как органам власти, отвечающим

за обеспечение информационной безопасности в стране, так и компаниям-поставщикам технологий, решений и сервисов в сфере информационной безопасности при определении направлений дальнейшего развития и использования цифровых технологий.

2 Концептуальная схема и показатели мониторинга доверия и безопасности при работе с большими данными

2.1 Обзор подходов к мониторингу и оценке уровня доверия и безопасности при работе с большими данными

Сегодня отсутствует общепринятая методология мониторинга и оценки уровня доверия и безопасности при работе с большими данными. Существующие подходы к такой деятельности в основном направлены на анализ обеспечения информационной безопасности и доверия применительно к использованию цифровых технологий в целом.

Прежде всего следует отметить деятельность по оценке и сопоставлению уровня развития информационной безопасности в разных странах, осуществляемую в рамках подготовки Глобального индекса кибербезопасности Международного союза электросвязи [8]. Данный индекс отражает уровень обеспечения информационной безопасности по странам, при этом оцениваются следующие виды деятельности по обеспечению кибербезопасности:

- правовые меры;
- технические меры;
- организационные меры;
- меры по развитию потенциала страны в сфере информационной безопасности и доверия;
- меры по организации сотрудничества в этой сфере (международное, межведомственное, межотраслевое).

Индекс публикуется раз в 2 года; в 2020 году Россия заняла 5-е место, набрав 98,06 баллов из 100.

Еще одним подходом к мониторингу и оценке уровня обеспечения кибербезопасности является Национальный индекс кибербезопасности [9] Академии электронного управления, в котором оценивается готовность стран к предотвращению киберугроз и управлению инцидентами, связанными с информационной безопасностью. В основу концептуальной схемы Национального индекса положены угрозы информационной безопасности, реагировать на которые обязана каждая страна: недоступность электронных сервисов; нарушение целостности данных и нарушение конфиденциальности данных. Индекс фокусируется на измеримых аспектах деятельности по обеспечению кибербезопасности среди которых:

- действующее законодательство в сфере кибербезопасности;
- наличие организационных структур, ответственных за обеспечение информационной безопасности;
- механизмы сотрудничества в сфере обеспечения кибербезопасности;
- конкретные результаты деятельности: стратегии и политики, технологические решения, планы развития.

По состоянию на 2020 год в Национальном индексе кибербезопасности Россия находится на 29-м месте из 160.

Всемирным банком в сотрудничестве с Институтом развития информационного общества была разработана методика оценки уровня развития цифровой экономики DECA (Digital Economy Country Assessment) [10], предназначенная для различных стран мира. Одним из факторов, существенно влияющих на развитие цифровой экономики, оцениваемых в рамках указанной методики, выделены доверие и безопасность. Концептуальная схема предметной области данного фактора в DECA включает следующие аспекты:

- государственная политика и регулирование (включает оценку национальной политики в сфере обеспечения информационной безопасности, а также наличие мероприятий, направленных на повышение осведомленности граждан и организаций в сфере информационной безопасности при использовании цифровых технологий);
- организационные меры по обеспечению информационной безопасности (содержит оценку групп реагирования на чрезвычайные ситуации в области информационной

безопасности, а также наличие механизмов государственно-частного партнерства и координации вопросов обеспечения информационной безопасности).

Среди имеющихся подходов к мониторингу и оценке информационной безопасности и доверия следует также отметить деятельность Организации по экономическому сотрудничеству и развитию, которая разработала методологии [11,12] и публикует данные [13] результатов опросов граждан и организаций по вопросам использования информационных технологий, включая аспекты, связанные с оценкой уровня информационной безопасности и доверия.

Несмотря на то, что описанные подходы напрямую не относятся к обеспечению информационной безопасности и доверия при работе с большими данными, основные аспекты приведенных методологий учитываются при формировании концептуальной схемы для целей настоящего исследования.

2.2 Обзор литературы по мониторингу и оценке уровня доверия и безопасности при работе с большими данными

Для разработки концептуальной схемы мониторинга и оценки уровня доверия и безопасности работы с большими данными авторами проведен анализ релевантных научных публикаций, индексируемых в платформе Web of Science (WoS).

С этой целью из «ядерной» коллекции WoS (WoS Core Collection) был выделен массив публикаций в области технологий работы с большими данными на основе следующего поискового образа:

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Детальное описание процедуры формирования приведенного выше поискового образа описано в статье [7]. За временной период с 2016 по 2020 год с использованием поиска по названию, аннотации, автору и ключевым словам (применено поле «Тема» в инструментарии InCites платформы WoS) было найдено 108 073 публикаций, проиндексированных в WoS.

На следующем этапе из этого массива осуществлялся отбор публикаций, имеющих отношение к доверию и безопасности, путем задания специализированного поискового образа:

cybersecurity OR “information security” OR trust OR “cybersecurity strategy*” OR “cybersecurity policy*” OR “personal data*” OR “organizational security” OR “technical security” OR “anonymization” OR “privacy by design”

Приведенные ключевые слова, с одной стороны, позволяют получить максимальное покрытие исследуемой области, а с другой – не дают значительного количества «шума», то есть публикаций, слабо связанных с темой исследования (для этого была проведена экспериментальная работа методом экспертных оценок с последующей верификацией по отсеиванию специализированных терминов из первоначального списка, который включал 31 наименование).

Общий и специализированный поисковые запросы были объединены логическим оператором «AND», и в результате после поиска по полю «Тема», было получено 2482 публикации за период с 2016 по 2020 год.

Следующим этапом по обзору публикаций был анализ полученных в результате поиска работ на предмет релевантности исследуемой теме. После анализа наименований, ключевых слов и аннотаций было отобрано 49 публикаций, для которых было принято решение об ознакомлении с полными текстами с целью выделения аспектов, относящихся к мониторингу и оценке доверия и безопасности при работе с большими данными.

Следует отметить, что среди отобранных публикаций встречаются те, что связаны с теоретическими и практическими подходами к использованию технологий работы с большими данными при обеспечении информационной безопасности традиционных информационных систем и цифровой инфраструктуры, однако данное направление не является предметом

настоящего исследования. Изучение полных текстов отобранных публикаций привело к отсеиванию еще 9 работ как нерелевантных.

Оставшиеся 40 публикаций [14-53] можно разделить на две группы:

- публикации, исследующие вопросы доверия и безопасности граждан при работе с большими данными (14 статей);
- работы, исследующие доверие и безопасность при работе с большими данными в организациях (26 статей).

Анализ публикаций из первой группы показал, что основными аспектами, связанными с доверием и безопасностью граждан, являются ограничения и меры защиты, принимаемые гражданами при использовании цифровых сервисов, предоставляемых компаниями, основная деятельность которых связана с обработкой больших данных (социальные сети, маркетплейсы, облачные хранилища, мессенджеры и т.д.). В частности, 10 публикаций в той или иной мере посвящены исследованию ограничений в доступе к функциональности смартфона, принимаемых пользователями при работе с приложениями [14-23]; 13 публикаций исследуют вопросы ограничения пользователей в доступе к своим данным, а также к данным о себе в социальных сетях и облачных хранилищах [14, 16-27]; 9 публикаций относятся к вопросам средств и методов защиты, принимаемых пользователями при работе с цифровыми платформами и сервисами [14, 16-18, 21-24,27].

Анализ публикаций из второй группы, посвященных исследованию доверия и безопасности при работе с большими данными в организациях, позволил выделить следующие аспекты, влияющие на успешность этой деятельности: организационные меры обеспечения информационной безопасности при работе с большими данными (политики, концепции, распорядительные акты, регулирующие информационную безопасность компании в целом, а также безопасность при работе с большими данными в частности) – 13 публикаций [28-40]; технические меры обеспечения информационной безопасности (традиционные [15, 32, 34, 36, 40-42] и специализированные, применяемые при работе с большими данными [13-16, 26, 28, 31-38, 40-42, 45-52]); наличие инцидентов информационной безопасности, связанных с раскрытием, повреждением или уничтожением данных организации – 6 публикаций [15, 27, 31, 35, 40, 53].

Проведенный анализ позволяет сделать вывод, что на сегодня не существует общепризнанных подходов к мониторингу и оценке уровня доверия и безопасности при работе с большими данными для отдельно взятой страны. В то же время явно выделяются два основных направления: исследования вопросов безопасности и доверия граждан при работе с большими данными, а также изучение процессов обеспечения информационной безопасности и доверия в организациях, использующих технологии работы с большими данными. Выделенные направления легли в основу разрабатываемой концептуальной схемы для мониторинга и оценки уровня доверия и безопасности при работе с большими данными.

2.3 Концептуальная схема мониторинга

С учетом проведенного анализа доступных источников, релевантных предмету исследования, была сформирована концептуальная схема мониторинга и оценки уровня доверия и безопасности работы с большими данными в России, которая представлена на рисунке 1.

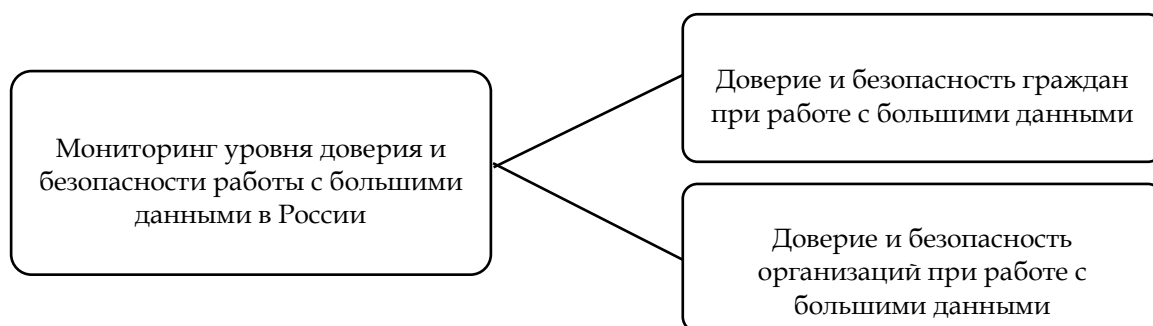


Рисунок 1. Концептуальная схема мониторинга уровня доверия и безопасности работы с большими данными в России

Концептуальная схема включает в себя два компонента: доверие и безопасность граждан при работе с большими данными и доверие и безопасность организаций при работе с большими данными).

2.4 Показатели доверия и безопасности граждан при работе с большими данными

Деятельность пользователей является важным источником больших данных. Множество существующих решений и сервисов основано на обработке и придании добавленной ценности данным, которые фактически производят пользователи в процессе использования приложений и сервисов. Граждане передают компаниям данные о своем местонахождении, выкладывают аудио-, фото- и видеоконтент, осуществляют взаимодействия с другими пользователями через социальные сети и с помощью мессенджеров, хранят свои данные в облачных хранилищах. Важным источником больших данных являются результаты деятельности граждан во Всемирной паутине, которые фиксируются в файлах «куки». Все эти «цифровые следы» наряду с данными, собираемыми поисковыми машинами, являются ценными для маркетинговых компаний, которые на их основе анализируют потребительские предпочтения и формируют персонализированные предложения, а также используют для развития своих продуктов и услуг.

В то же время, полномасштабное развитие технологий и сервисов, основанных на обработке больших массивов пользовательских данных, влечет за собой риски деанонимизации пользователей и некорректного использования их персональных данных, что приводит к потенциально высокому уровню недоверия к использованию такого рода сервисов. В условиях, когда большинство граждан ограничивают доступ к своим данным, развитие сервисов и технологий, основанных на работе с большими данными, крайне затруднительно.

Таким образом, при мониторинге и оценке доверия и безопасности граждан при работе с большими данными предлагается оценить долю граждан, открывающих доступ к своим персональным данным, для чего будут использованы следующие показатели:

- (ДБГ-01-01) Доля граждан, ограничивавших доступ приложениям на своем смартфоне к местоположению, списку контактов, фото/видео контенту, камере, микрофону или отказывавших в таком доступе;
- (ДБГ-01-02) Доля граждан, ограничивавших доступ к сведениям о себе и к своим данным, хранящимся в социальных сетях и (или) в совместно используемых облачных хранилищах.

В рамках оценки уровня доверия и безопасности при работе с большими данными в России видится целесообразным изучить уровень компетенций граждан по защите и/или управлению доступом к своим персональным данным. Чем выше компетенции граждан, тем с большим доверием они будут относиться к использованию сервисов, основанных на больших данных. Мониторинг данного направления позволит оценить, с одной стороны, уровень компетенций граждан по управлению доступом к своим данным, а с другой – уровень доверия пользователей к сервисам, собирающим и использующим их данные.

Для оценки уровня компетенций граждан при управлении доступом к своим данным предлагается использовать следующие показатели:

- (ДБГ-01-03) Доля граждан, изменявших настройки в своем веб-браузере, чтобы предотвратить или ограничить использование файлов «куки»;
- (ДБГ-01-04) Доля граждан, использовавших программное обеспечение, которое ограничивает возможность отслеживания деятельности в интернете.

2.5 Показатели доверия и безопасности при работе с большими данными в организации

Еще одной составляющей данной предметной области мониторинга и оценки является уровень доверия и безопасности при работе с большими данными в организациях. Деятельность по обеспечению информационной безопасности является исключительно важной для всех типов организаций, работающих с большими данными, включая органы государственной власти и местного самоуправления, частные и государственные компании, медицинские и образовательные организации.

Как отмечалось в разделе 1, мониторинг и оценку уровня доверия и безопасности при работе с большими данными в организациях планируется проводить в 10 сферах деятельности:

- система государственного управления;
- здравоохранение;

- образование
- наука;
- промышленность;
- сельское хозяйство;
- строительство;
- развитие городской среды;
- транспорт и логистика;
- энергетическая инфраструктура;
- финансовые услуги.

Основным источником данных для получения значений показателей и анализа ситуации в различных сферах деятельности являются выборочные представительные опросы организаций. Более подробно методология сбора данных и вычисления значений показателей по отдельным предметным областям мониторинга и оценки описана в статье [13].

Возможности технологий работы с большими данными порождают множество рисков информационной безопасности в организациях. Инструменты работы с ними являются зачастую инновационными, что потенциально может привести к невозможности прогнозирования всех потенциальных угроз. Традиционные средства обеспечения информационной безопасности могут не сработать в условиях больших массивов разнообразных данных и значительного количества источников, а также в случае высокой скорости генерации потоков данных. Распределенные хранилища и множество вовлеченного персонала в свою очередь порождают дополнительные риски умышленных или непреднамеренных утечек данных, а также возможность их незапланированного раскрытия.

Исходя из проведенного анализа (см. раздел 2) были выделены три направления мониторинга и оценки уровня безопасности и доверия при работе с большими данными в организации: организационные меры, технические меры, а также частота возникновения в организациях инцидентов, связанных с компрометацией, повреждением или уничтожением данных.

При мониторинге организационных мер обеспечения доверия и безопасности используются два показателя, связанные с наличием в организации политики обеспечения безопасности и доверия – как общей, так и специализированной, направленной на применение технологий работы с большими данными:

- (ДБО-02-01) Доля организаций, имеющих политику обеспечения информационной безопасности;
- (ДБО-02-02) Доля организаций, имеющих специализированную политику обеспечения информационной безопасности для работы с большими данными.

В соответствии с национальным стандартом ГОСТ Р 53114–2008 [54], под политикой (обеспечения) информационной безопасности понимается формальное изложение правил поведения, процедур, практических приемов или руководящих принципов в области информационной безопасности, которыми руководствуется организация в своей деятельности.

Политика обеспечения информационной безопасности при работе с большими данными должна учитывать основные характеристики работы с большими данными, которые порождают дополнительные риски, не учитываемые ранее. Сверхбольшие объемы данных требуют обеспечения информационной безопасности многоуровневых распределенных хранилищ; высокая скорость генерации больших потоков данных требует защиты процессов распределенных вычислений, мониторинга инцидентов в режиме реального времени; однако применение традиционных средств защиты может существенно снизить скорость передачи и обработки данных. Разнообразные форматы структурированных и неструктурированных данных (текст, аудио, видео, изображения) и множество источников их генерации приводят к необходимости применения различных средств защиты, которые должны работать в рамках распределенной инфраструктуры как единое целое.

Политика обеспечения информационной безопасности работы с большими данными может быть утверждена как отдельно, так и быть частью общей политики обеспечения информационной безопасности организации.

Вторым направлением мониторинга и оценки уровня доверия и безопасности при работе с большими данными в организации является оценка интенсивности применения мер обеспечения информационной безопасности в этой организации. В рамках разработанной концептуальной

схемы меры обеспечения информационной безопасности разделяются на базовые (требуемые для обеспечения необходимого уровня информационной защищенности цифровых технологий, используемых организацией) и специализированные (направленные на обеспечение информационной безопасности технологий работы с большими данными). Для этого в концептуальную схему мониторинга BD4DE включены следующие показатели:

- (ДБО-02-03) Доля организаций, применяющих меры обеспечения информационной безопасности при работе с цифровыми технологиями;
- (ДБО-02-04) Доля организаций, применяющих специализированные меры обеспечения информационной безопасности для работы с большими данными.

Под мерой обеспечения информационной безопасности в соответствии с упомянутым выше стандартом [54] понимается совокупность действий, направленных на разработку и/или практическое применение способов и средств обеспечения информационной безопасности. Для оценки показателя ДБО-02-03 была выбрана совокупность мер, используемых организациями, которые отслеживаются Росстатом в рамках федерального статистического наблюдения за использованием информационно-коммуникационных технологий [5].

Еще одним направлением мониторинга в данной предметной области является оценка частоты возникновения инцидентов информационной безопасности, связанных с повреждением или уничтожением данных, а также приведших к утечке конфиденциальных данных. Для оценки этих аспектов в концептуальной схеме BD4DE используются следующие показатели:

- (ДБО-02-05) Доля организаций, столкнувшихся с инцидентами информационной безопасности, связанными с уничтожением или повреждением данных;
- (ДБО-02-06) Доля организаций, столкнувшихся с инцидентами информационной безопасности, связанными с раскрытием конфиденциальных данных.

К инцидентам информационной безопасности, приведшим к уничтожению или повреждению данных, могут относиться заражения вредоносным программным обеспечением, несанкционированные проникновения, сбои аппаратного или программного обеспечения.

К инцидентам информационной безопасности, приведшим к утечке конфиденциальных данных, могут относиться несанкционированные проникновения, фарминг (скрытное перенаправление жертвы на ложный IP-адрес), фишинг-атаки (обман пользователей с целью раскрытия конфиденциальных данных), действия собственных сотрудников (преднамеренные или непреднамеренные).

Большое количество инцидентов может влиять на уровень доверия к технологиям работы с большими данными, а также сигнализировать о недостаточном уровне развития организационных и технических мер, принимаемых организацией для обеспечения доверия и безопасности.

3 Методология измерения показателей доверия и безопасности работы с большими данными

Источником данных для расчета значений показателя ДБО-02-03 являются итоги федерального статистического наблюдения за использованием информационно-коммуникационных технологий в организациях.

Для показателей системы мониторинга BD4DE, которые не отслеживаются Федеральной службой государственной статистики, требуется проводить опрос граждан и организаций, для чего разработан соответствующий инструментарий (анкета, паспорта показателей).

Анкета для опроса организаций содержит блок модельных вопросов, позволяющих получить сведения о значениях каждого показателя, предусмотренного для мониторинга и оценки. Анкета содержит определения специализированных терминов, а также пояснения, необходимые для однозначной трактовки задаваемых респондентам вопросов.

Для оценки уровня использования технических мер обеспечения информационной безопасности при работе организаций с цифровыми технологиями (показатель ДБО-02-03) выбраны следующие меры:

- средства строгой аутентификации (например, пароли длиннее 8 символов, сменяемые не реже чем раз в 6 месяцев);
- технические средства аутентификации пользователей (например, токены, USB-ключи или смарт-карты);

- резервное копирование данных на носители, находящиеся физически не на территории организации;
- биометрические средства аутентификации пользователей;
- средства шифрования;
- средства электронной подписи;
- регулярно обновляемые антивирусные программы;
- программные / аппаратные средства, препятствующие несанкционированному доступу вредоносных программ из глобальных информационных / локальных вычислительных сетей (брандмауэр);
- спам-фильтр;
- системы обнаружения вторжения в компьютер или сеть;
- программные средства автоматизации процессов анализа и контроля защищенности компьютерных систем.

Для оценки уровня использования специализированных технических мер защиты, направленных на обеспечение доверия и безопасности при работе с большими данными в организациях (см. показатель ДБО-02-04), выбраны следующие меры:

- средства защиты нереляционных хранилищ данных (например, специализированные алгоритмы криптографического хеширования, такие как SHA-256 или выше, использование протоколов TLS / SSL);
- средства защиты многоуровневых хранилищ данных (например, SUNDR, криптографические облачные хранилища данных);
- криптографические средства обеспечения контроля доступа к данным (например, гомоморфное шифрование, реляционное шифрование, системы, основанные на идентификаторах);
- меры защиты инструментов распределенных вычислений (например, использование протокола Kerberos, технологии мандатного управление доступом);
- средства мониторинга инцидентов и аудита системы обеспечения информационной безопасности в режиме реального времени (например, SIEM решения, средства мониторинга и аудита на всех стадиях жизненного цикла данных);
- средства фильтрации и валидации данных, собираемых из распределенных источников (средства тестирования инфраструктурных ресурсов, использование доверенных сертификатов, использование доверенных устройств и приложений, использование на точках сбора данных средств защиты от вредоносных программ, метрики доверия к узлам и поставщикам данных);
- средства предотвращения деанонимизации пользовательских данных (например, использование технологий дифференциальной конфиденциальности, гомоморфного шифрования);
- другие меры.

4 Результаты пилотной реализации

На первой стадии пилотной реализации было принято решение расширить область мониторинга деятельности организаций показателями, характеризующими уровень развития человеческого капитала организации с точки зрения информационной безопасности. Было предложено дополнить концептуальную схему для данной предметной области показателем, который характеризует наличие у сотрудников организаций навыков по обеспечению доверия и безопасности при работе с большими данными. Другим важным фактором, влияющим на работу с большими данными, является обеспеченность организации специалистами по информационной безопасности, которые способны отвечать на имеющиеся вызовы в сфере информационной безопасности и использовать специализированные инструменты защиты при работе с большими данными.

По результатам первой стадии пилотной реализации были доработаны соответствующие разделы анкет для обследования организаций по вопросам использования технологий работы с большими данными.

На второй стадии пилотной реализации системы мониторинга проводилась экспертная оценка разработанных анкет экспертом данной предметной области (специалистом по

информационной безопасности) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций).

Экспертом по информационной безопасности были сформулированы следующие предложения:

- дополнить анкету отдельным вопросом о наличии в организации концепции информационной безопасности, включающей разделы, посвященные безопасности при работе с большими данными;
- дополнить перечень мер обеспечения информационной безопасности при использовании цифровых технологий средствами защиты облачных вычислений;
- дополнить перечень мер обеспечения информационной безопасности при работе с большими данными средствами защиты, предоставляемыми сторонними организациями как услуги («Security-as-a-Service» – служба облачных вычислений, обеспечивающая информационную безопасность).

Кроме того, экспертом по информационной безопасности было высказано предположение, что на вопросы анкеты, связанные с обеспеченностью кадрами, большинством респондентов будут даны отрицательные ответы.

Экспертом-социологом были сформулированы следующие предложения: дать пояснения к разделам анкеты о том, какого уровня (в иерархии организации) и какой квалификации специалистов требуется привлечь в качестве респондентов для ответов на вопросы соответствующих разделов; уточнить понятийный аппарат и комментарии по отдельным вопросам.

На последней стадии пилотной реализации полученные замечания и предложения экспертов были учтены в соответствующих разделах финальных версий анкет для опроса организаций.

В настоящее время доступным для расчета показателем, основанным на данных Росстата, является уровень использования технических мер обеспечения информационной безопасности при работе с цифровыми технологиями (показатель ДБО-02-03). Результаты расчета значений показателя по итогам 2019 года [5] для выбранных отраслей представлены ниже на рисунках 2–3.

На рисунке 2 показана доля организаций в различных сферах деятельности, подтвердивших использование хотя бы одной из следующих технических мер при использовании цифровых технологий: средства строгой аутентификации; технические средства аутентификации пользователей; резервное копирование данных на носители, находящиеся физически не на территории организации; биометрические средства аутентификации пользователей.

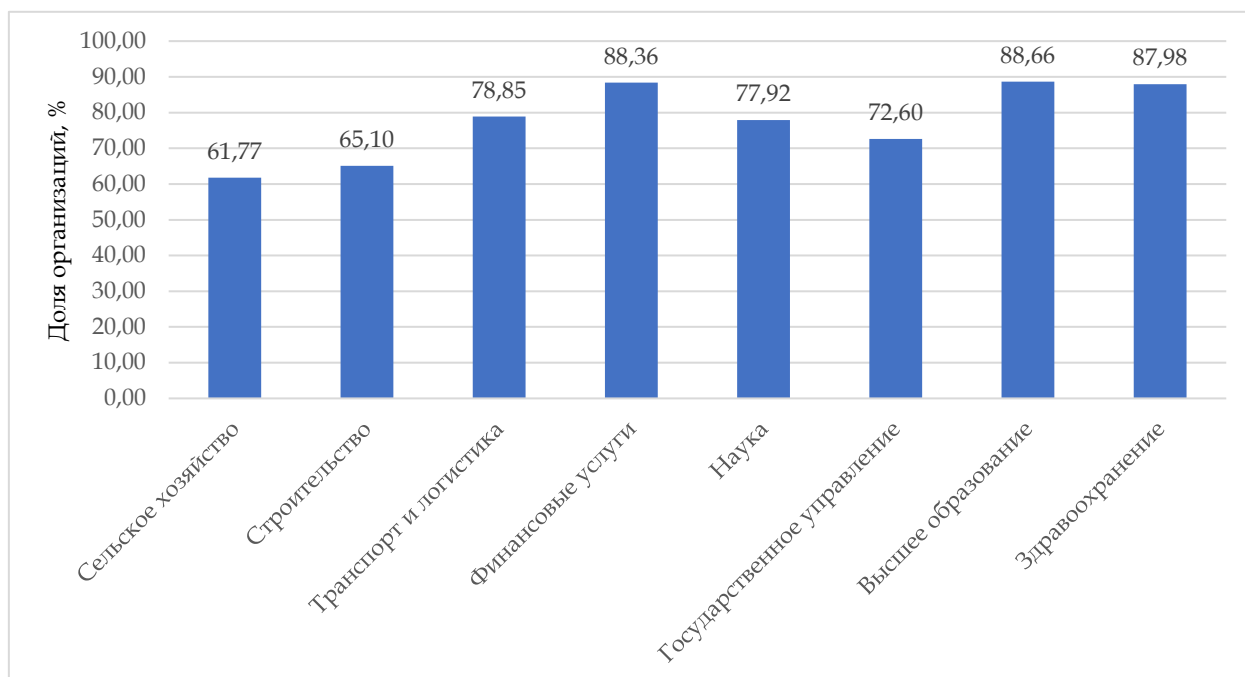


Рисунок 2. Доля российских организаций, использующих средства обеспечения информационной безопасности (в разрезе приоритетных сфер деятельности)

Наиболее популярной мерой из вышеприведенного списка являются технические средства аутентификации пользователей (например, токены, USB-ключи или смарт-карты). Их используют более 50% организаций во всех рассматриваемых сферах деятельности (лидер – высшее образование, 79,43%). Наименее популярной мерой являются биометрические средства аутентификации пользователей. Во всех сферах деятельности, кроме финансовых услуг (24%) уровень использования не превышает 6%.

На рисунке 3 представлены организации в различных сферах деятельности, подтвердившие использование хотя бы одной из следующих технических мер при использовании цифровых технологий:

- средства шифрования;
- средства электронной подписи;
- регулярно обновляемые антивирусные программы;
- программные / аппаратные средства, препятствующие несанкционированному доступу вредоносных программ из глобальных информационных / локальных вычислительных сетей (брандмауэр);
- спам-фильтры;
- системы обнаружения вторжения в компьютер или сеть;
- программные средства автоматизации процессов анализа и контроля защищенности компьютерных систем.

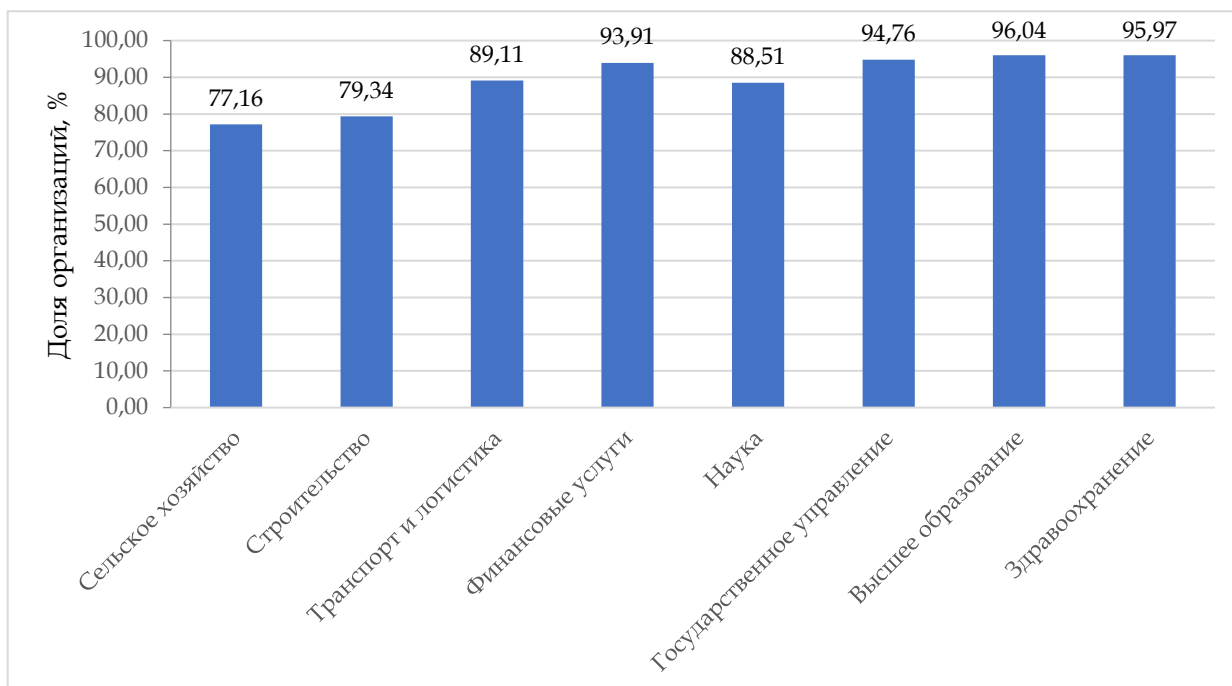


Рисунок 3. Доля российских организаций, использующих средства защиты информации, передаваемой по глобальным сетям (в разрезе приоритетных сфер деятельности)

Наиболее распространенной мерой обеспечения информационной безопасности из приведенного выше списка являются средства электронной подписи, их использование во всех рассматриваемых сферах деятельности находится на уровне 69% и выше. Меньше всего распространены программные средства автоматизации процессов анализа и контроля защищенности компьютерных систем, за исключением финансовой отрасли, в которой данную меру применяют более 60% организаций, и сферы высшего образования (47%); в исследуемых сферах деятельности уровень использования указанного средства обеспечения информационной безопасности не превышает 37%.

Следует отметить, что на графиках не представлены результаты для таких сфер деятельности, как обрабатывающая промышленность, развитие городской среды и энергетическая инфраструктура. Для получения данных по этим отраслям необходимо делать специализированный запрос данных в Главный межрегиональный центр обработки и

распространения статистической информации Росстата России, содержащий сложные группировки кодов ОКВЭД 2, описывающих совокупность видов деятельности, входящих в указанные сферы деятельности, что видится нецелесообразным в рамках пилотной реализации разработанной концептуальной схемы. Сфера деятельности «Образование» ограничена высшим образованием, так как сведения по организациям среднего общего и профессионального образования Росстатом не собираются. Наличие двух графиков для показателя ДБО-02-03 обусловлено спецификой сбора и публикации данных Росстатом, который разделяет технические меры обеспечения информационной безопасности при работе организаций с цифровыми технологиями на два списка.

Понятно, что в полной мере оценить актуальный уровень доверия и безопасности работы с большими данными в России только по указанному показателю не представляется возможным. Данные, необходимые для расчета отобранных показателей, требуется получать в рамках представительного опроса граждан и организаций.

Заключение

В данной работе описана методология мониторинга и оценки уровня доверия и безопасности работы с большими данными в России. Концептуальная схема, предлагаемая для использования в рамках методологии, основывается на изучении релевантных научных публикаций, выведших за последние 5 лет, а также работ ведущих аналитических и исследовательских международных организаций.

Применимость предлагаемой концептуальной схемы и вошедших в нее показателей мониторинга доказана в рамках пилотной реализации с учетом рекомендованных доработок. Следующим этапом апробации предложенного подхода должно стать масштабное полевое исследование, включающее в себя опрос граждан, а также организаций из приоритетных отраслей экономики и социальной сферы с использованием разработанных анкет. В настоящее время данные для расчета значений большей части показателей недоступны.

Дальнейшее развитие настоящего исследования может идти в направлении уточнения концептуальной схемы мониторинга, а также пересмотре перечня показателей, который может потребоваться после сбора всех необходимых данных для проведения оценки, а также получения обратной связи от опрашиваемых организаций и профессионального сообщества.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ имени М.В.Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

Литература

1. Стратегия национальной безопасности Российской Федерации. Утверждена Указом Президента Российской Федерации от 2 июля 2021 года № 400
2. Доктрина информационной безопасности Российской Федерации. Утверждена Указом Президента Российской Федерации от 5 декабря 2016 г. № 646
3. Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы. Утверждена Указом Президента Российской Федерации от 9 мая 2017 года № 203
4. Национальный проект «Цифровая экономика». Утвержден на заседании президиума Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам 24 декабря 2018 года.
5. Федеральная служба государственной статистики. Итоги федерального статистического наблюдения по ф. № 3-информ «Сведения об использовании информационных и

- коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказании услуг в этих сферах». URL: <https://rosstat.gov.ru/folder/14478> (дата обращения 01.06.2021)
6. Федеральная служба государственной статистики. Итоги федерального статистического наблюдения по ф. № 1-ИТ «Анкета выборочного федерального статистического наблюдения по вопросам использования населением информационных технологий и информационно-телекоммуникационных сетей. URL: https://gks.ru/free_doc/new_site/business/it/ikt20/index.html (дата обращения 01.06.2021)
 7. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
 8. International Telecommunication Union Development Sector. Global Cybersecurity Index 2020. URL : https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2021-PDF-E.pdf (дата обращения 01.06.2021)
 9. e-Governance Academy Foundation. The National Cyber Security Index. URL: <https://ncsi.ega.ee/methodology/> (дата обращения 01.06.2021)
 10. Институт развития информационного общества. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. – 166 с
 11. The OECD Model Survey on ICT Access and Usage by Households and Individuals. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Access-Usage-Households-Individuals.pdf> (дата обращения 01.06.2021)
 12. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (дата обращения 01.06.2021)
 13. OECD. Data. URL: <https://stats.oecd.org/> (дата обращения 01.06.2021)
 14. Yuxue Li Lijun Song Yucheng Zeng. Research on information security and privacy protection model based on consumer behavior in big data environment. June 2018. <https://doi.org/10.1002/cpe.4881>
 15. Nils Gruschka, Vasileios Mavroedisy, Kamer Vishiy, Meiko Jensen. Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. 2018 IEEE International Conference on Big Data (Big Data). <https://doi.org/10.1109/BigData.2018.8622621>
 16. Luca Bolognini, Camilla Bistolfi, Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), <https://doi.org/10.1016/j.clsr.2016.11.002>
 17. Grout, V. No More Privacy Any More? Information 2019, 10, 19. <https://doi.org/10.3390/info10010019>
 18. Carole L. Jurkiewicz (2018): Big Data, Big Concerns: Ethics in the Digital Age, Public Integrity, <https://doi.org/10.1080/10999922.2018.1448218>
 19. Alessandro Mantelero, Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework, Computer Law & Security Review: The International Journal of Technology Law and Practice (2017), <https://doi.org/10.1016/j.clsr.2017.05.011>
 20. Martin, K.D., Murphy, P.E. The role of data privacy in marketing. J. of the Acad. Mark. Sci. 45, 135–155 (2017). <https://doi.org/10.1007/s11747-016-0495-4>
 21. Muhammad, S.S., Dey, B.L. & Weerakkody, V. Analysis of Factors that Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature. Inf Syst Front 20, 559–576 (2018). <https://doi.org/10.1007/s10796-017-9802-y>
 22. N. A. Shози and J. Mtsweni, "Big data privacy in social media sites," 2017 IST-Africa Week Conference (IST-Africa), 2017, pp. 1-6, <https://doi.org/10.23919/ISTAfrICA.2017.8102311>.
 23. L. Yuqing, "Research on Personal Information Security on Social Network in Big Data Era," 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), 2017, pp. 676-678, <https://doi.org/10.1109/ICSGEA.2017.91>.
 24. Liu H (2018) Beyond the Scale of Big Data. Front. Big Data 1:1. doi: 10.3389/fdata.2018.00001
 25. E. Kosta et al. (Eds.): Privacy and Identity 2018, IFIP AICT 547, pp. 81–94, 2019. https://doi.org/10.1007/978-3-030-16744-8_6

26. Fan, K., Lou, S., Su, R. et al. Secure and private key management scheme in big data networking. Peer-to-Peer Netw. Appl. 11, 992-999 (2018). <https://doi.org/10.1007/s12083-017-0579-z>
27. SanchengPeng, ShuiYu, PeterMueller. Social networking big data: Opportunities, solutions, and challenges. Future Generation Computer Systems 86 (2018) 1456-1458. doi: <https://doi.org/10.1016/j.future.2018.05.040>
28. M. Binjubeir, A. A. Ahmed, M. A. B. Ismail, A. S. Sadiq and M. Khurram Khan, "Comprehensive Survey on Big Data Privacy Protection," in IEEE Access, vol. 8, pp. 20067-20079, 2020, <https://doi.org/10.1109/ACCESS.2019.2962368>.
29. Poltavtseva, M.A. A Consistent Approach to Building Secure Big Data Processing and Storage Systems. Aut. Control Comp. Sci. 53, 914-921 (2019). <https://doi.org/10.3103/S0146411619080273>
30. Poltavtseva, M.A., Kalinin, M.O. Modeling Big Data Management Systems in Information Security. Aut. Control Comp. Sci. 53, 895-902 (2019). <https://doi.org/10.3103/S014641161908025X>
31. Hazirah Bee Yusof Ali, Lili Marziana bt Abdullah, Mira Kartiwi, Azlin Nordin. Risk Assessment for Big Data in Cloud: Security, Privacy and Trust. the 2018 Artificial Intelligence and Cloud Computing Conference. <https://doi.org/10.1145/3299819.3299841>
32. Chang, Y., Government Information Quarterly (2018), <https://doi.org/10.1016/j.giq.2018.04.002>
33. El Haourani L., Abou El Kalam A., Ait Ouahman A. (2019) Knowledge Based Access Control a Model for Security and Privacy in the Big Data. In: Ben Ahmed M., Boudhir A., Younes A. (eds) Innovations in Smart Cities Applications Edition 2. SCA 2018. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham. https://doi.org/10.1007/978-3-030-11196-0_60
34. N. Miloslavskaya, A. Nikiforov and V. Budzko, Standardization of Ensuring Information Security for Big Data Technologies, 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), 2018, pp. 56-63, <https://doi.org/10.1109/W-FiCloud.2018.00015>
35. Reza Saneei Moghadam, Ricardo Colomo-Palacios. Information security governance in big data environments: A systematic mapping, Procedia Computer Science, Volume 138, 2018, Pages 401-408, <https://doi.org/10.1016/j.procs.2018.10.057>
36. Khairulliza Ahmad Salleh, Lech Janczewski, Security Considerations in Big Data Solutions Adoption: Lessons from a Case Study on a Banking Institution, Procedia Computer Science, Volume 164, 2019, Pages 168-176, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.12.169>
37. Shetty M.M., Manjaiah D.H., Hemdan E.ED. (2019) Policy-Based Access Control Scheme for Securing Hadoop Ecosystem. In: Balas V., Sharma N., Chakrabarti A. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing, vol 839. Springer, Singapore. https://doi.org/10.1007/978-981-13-1274-8_13
38. M. Tang, M. Alazab and Y. Luo, "Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies," in IEEE Transactions on Big Data, vol. 5, no. 3, pp. 317-329, 1 Sept. 2019, <https://doi.org/10.1109/TBDATA.2017.2723570>
39. F. Khafa et al. (eds.), Research into Information Security Strategy. Practices for Commercial Banks in Taiwan Recent Developments in Intelligent Systems and Interactive Applications, Advances in Intelligent Systems and Computing 541, https://doi.org/10.1007/978-3-319-49568-2_25
40. A. Yarali, R. Joyce and B. Dixon, "Ethics of Big Data: Privacy, Security and Trust," 2020 Wireless Telecommunications Symposium (WTS), 2020, pp. 1-7, <https://doi.org/10.1109/WTS48268.2020.9198734>
41. Sahel Alouneh, Ismail Hababeh, Tamer Alajrami. Toward big data analysis to improve enterprise information security. MEDES '18: Proceedings of the 10th International Conference on Management of Digital EcoSystems September 2018 Pages 106-109 <https://doi.org/10.1145/3281375.3281393>
42. Macklin, Thomas ; Mathews, Joseph. Big data, little security: Addressing security issues in your platform. Proceedings of the SPIE, Volume 10207, id. 102070G 10 pp. (2017). <https://doi.org/10.1117/12.2268002>
43. Junjun Guo, Le Wang, Learning to upgrade internet information security and protection strategy in big data era, Computer Communications, Volume 160, 2020, Pages 150-157, <https://doi.org/10.1016/j.comcom.2020.05.043>
44. Maanak Gupta, Farhan Patwa, Ravi Sandhu . An Attribute-Based Access Control Model for

45. Secure Big Data Processing in Hadoop Ecosystem. the Third ACM Workshop. <https://doi.org/10.1145/3180457.3180463>
46. Colombo, P., Ferrari, E. Access control technologies for Big Data management systems: literature review and future trends. *Cybersecur* 2, 3 (2019). <https://doi.org/10.1186/s42400-018-0020-9>
47. N. Lee and B. Wu, "Privacy Protection Technology and Access Control Mechanism for Medical Big Data," 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017, pp. 424-429, <https://doi.org/10.1109/IIAI-AAI.2017.34>
48. D. Mondek, R. B. Blažek and T. Zahradnický, "Security Analytics in the Big Data Era," 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2017, pp. 605-606, <https://doi.org/10.1109/QRS-C.2017.136>
49. Meisuchi Naisuty, Achmad Nizar Hidayanto, Nabila Clydea Harahap, Ahmad Rosyiq, Agus Suhanto and George Michael Samuel Hartono. Data protection on hadoop distributed file system by using encryption algorithms: a systematic literature review. 2020 J. Phys.: Conf. Ser. 1444 012012. <https://doi.org/10.1088/1742-6596/1444/1/012012>
50. Salas, J., Domingo-Ferrer, J. Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges. *Math.Comput.Sci.* 12, 263–274 (2018). <https://doi.org/10.1007/s11786-018-0344-6>
51. Hai Tao, Md Zakirul Alam Bhuiyan, Md Arafatur Rahman, Guojun Wang, Tian Wang, Md. Manjur Ahmed, Jing Li, Economic perspective analysis of protecting big data security and privacy, *Future Generation Computer Systems*, Volume 98, 2019, Pages 660-671, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2019.03.042>
52. Linqi Zhou, Weihong Gu, Cheng Huang, Aijun Huang, and Yongbin Bai. Research on information security in big data era. *AIP Conference Proceedings* 1967, 020020 (2018). <https://doi.org/10.1063/1.5038992>
53. Poltavtseva, M.A., Zegzhda, D.P. & Kalinin, M.O. Big Data Management System Security Threat Model. *Aut. Control Comp. Sci.* 53, 903–913 (2019). <https://doi.org/10.3103/S0146411619080261>
54. ГОСТ Р 53114–2008. Защита информации. Обеспечение информационной безопасности в организации. Основные термины и определения. Утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 18 декабря 2008 г. N 532-ст

TRUST AND SECURITY IN THE SPHERE OF WORKING WITH BIG DATA IN RUSSIA

Katin, Alexander Vladimirovich

Institute of the Information Society, head of Directorate of sectoral programs

Plekhanov Russian University of Economics, IIS-based Digital economy department, senior lecturer

Moscow, Russian Federation

alexander.katin@iis.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

Institute of the Information Society, chairman of the Board of directors

Plekhanov Russian University of Economics, IIS-Based Digital Economy Department, scientific advisor

Moscow, Russian Federation

yuri.hohlov@iis.ru

Abstract

The issues of monitoring and assessing the level of trust and security in working with big data in Russia have been investigated. A conceptual framework and a system of indicators have been developed for monitoring measures taken by citizens and organizations to ensure trust and security in working with big data. To assess the applicability of the developed framework, a pilot implementation was carried out.

Keywords

big data, big data technologies, information security, trust

References

1. Strategiya nacional'noj bezopasnosti Rossijskoj Federacii. Utverzhdena Ukazom Prezidenta Rossijskoj Federacii ot 2 iyulya 2021 goda № 400
2. Doktrina informacionnoj bezopasnosti Rossijskoj Federacii. Utverzhdena Ukazom Prezidenta Rossijskoj Federacii ot 5 dekabrya 2016 g. № 646
3. Strategiya razvitiya informacionnogo obshhestva v Rossijskoj Federacii na 2017- 2030 gody`. Utverzhdena Ukazom Prezidenta Rossijskoj Federacii ot 9 maya 2017 goda № 203
4. Nacional'ny`j proekt «Cifrovaya e`konomika». Utverzhden na zasedanii prezidiuma Soveta pri Prezidente Rossijskoj Federacii po strategicheskomu razvitiyu i nacional'ny`m proektam 24 dekabrya 2018 goda.
5. Federal'naya sluzhba gosudarstvennoj statistiki. Itogi federal'nogo statisticheskogo nablyudeniya po f. № 3-inform «Svedeniya ob ispol'zovanii informacionny`x i kommunikacionny`x texnologij i proizvodstve vy`chislitel'noj texniki, programmnogo obespecheniya i okazanii uslug v e`tix sferax». URL: <https://rosstat.gov.ru/folder/14478> (accessed on 01.06.2021).
6. Federal'naya sluzhba gosudarstvennoj statistiki. Itogi federal'nogo statisticheskogo nablyudeniya po f. № 1-IT «Anketa vy`borochnogo federal'nogo statisticheskogo nablyudeniya po voprosam ispol'zovaniya naseleniem informacionny`x texnologij i informacionno-telekommunikacionny`x setej. URL: https://gks.ru/free_doc/new_site/business/it/ikt20/index.html (accessed on 01.06.2021)
7. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya texnologij raboty` s bol'shimi dannymi // Informacionnoe obshhestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
8. International Telecommunication Union Development Sector. Global Cybersecurity Index 2020. URL: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2021-PDF-E.pdf (accessed on 01.06.2021).
9. e-Governance Academy Foundation. The National Cyber Security Index. URL: <https://ncsi.ega.ee/methodology/> (accessed on 01.06.2021)/

10. Institut razvitiya informacionnogo obshhestva. Analiz tekushhego sostoyaniya razvitiya cifrovoj ekonomiki v Rossii. M.: Institut razvitiya informacionnogo obshhestva, 2018. – 166 s.
11. The OECD Model Survey on ICT Access and Usage by Households and Individuals. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Access-Usage-Households-Individuals.pdf> (accessed on 01.06.2021)/
12. The OECD Model Survey on ICT Usage by Businesses. 2nd Revision. URL: <https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf> (accessed on 01.06.2021).
13. OECD. Data. URL: <https://stats.oecd.org/> (accessed on 01.06.2021)/
14. Yuxue Li Lijun Song Yucheng Zeng. Research on information security and privacy protection model based on consumer behavior in big data environment. June 2018. <https://doi.org/10.1002/cpe.4881>
15. Nils Gruschky, Vasileios Mavroeidisy, Kamer Vishiy, Meiko Jensen. Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. 2018 IEEE International Conference on Big Data (Big Data). <https://doi.org/10.1109/BigData.2018.8622621>
16. Luca Bolognini, Camilla Bistolfi, Pseudonymization and impacts of Big (personal/ anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), <https://doi.org/10.1016/j.clsr.2016.11.002>
17. Grout, V. No More Privacy Any More? Information 2019, 10, 19. <https://doi.org/10.3390/info10010019>
18. Carole L. Jurkiewicz (2018): Big Data, Big Concerns: Ethics in the Digital Age, Public Integrity, <https://doi.org/10.1080/10999922.2018.1448218>
19. Alessandro Mantelero, Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework, Computer Law & Security Review: The International Journal of Technology Law and Practice (2017), <https://doi.org/10.1016/j.clsr.2017.05.011>
20. Martin, K.D., Murphy, P.E. The role of data privacy in marketing. J. of the Acad. Mark. Sci. 45, 135–155 (2017). <https://doi.org/10.1007/s11747-016-0495-4>
21. Muhammad, S.S., Dey, B.L. & Weerakkody, V. Analysis of Factors that Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature. Inf Syst Front 20, 559–576 (2018). <https://doi.org/10.1007/s10796-017-9802-y>
22. N. A. Shozhi and J. Mtsweni, Big data privacy in social media sites, 2017 IST-Africa Week Conference (IST-Africa), 2017, pp. 1-6, <https://doi.org/10.23919/ISTAFRICA.2017.8102311>.
23. L. Yuqing, Research on Personal Information Security on Social Network in Big Data Era, 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), 2017, pp. 676-678, <https://doi.org/10.1109/ICSGEA.2017.91>
24. Liu H (2018) Beyond the Scale of Big Data. Front. Big Data 1:1. doi: 10.3389/fdata.2018.00001
25. E. Kosta et al. (Eds.): Privacy and Identity 2018, IFIP AICT 547, pp. 81–94, 2019. https://doi.org/10.1007/978-3-030-16744-8_6
26. Fan, K., Lou, S., Su, R. et al. Secure and private key management scheme in big data networking. Peer-to-Peer Netw. Appl. 11, 992–999 (2018). <https://doi.org/10.1007/s12083-017-0579-z>
27. SanchengPeng, ShuiYu, PeterMueller. Social networking big data: Opportunities, solutions, and challenges. Future Generation Computer Systems 86 (2018) 1456–1458. doi: <https://doi.org/10.1016/j.future.2018.05.040>
28. M. Binjubeir, A. A. Ahmed, M. A. B. Ismail, A. S. Sadiq and M. Khurram Khan, Comprehensive Survey on Big Data Privacy Protection, in IEEE Access, vol. 8, pp. 20067-20079, 2020, <https://doi.org/10.1109/ACCESS.2019.2962368>
29. Poltavtseva, M.A. A Consistent Approach to Building Secure Big Data Processing and Storage Systems. Aut. Control Comp. Sci. 53, 914–921 (2019). <https://doi.org/10.3103/S0146411619080273>
30. Poltavtseva, M.A., Kalinin, M.O. Modeling Big Data Management Systems in Information Security. Aut. Control Comp. Sci. 53, 895–902 (2019). <https://doi.org/10.3103/S014641161908025X>
31. Hazirah Bee Yusof Ali, Lili Marziana bt Abdullah, Mira Kartiwi, Azlin Nordin. Risk Assessment for Big Data in Cloud: Security, Privacy and Trust. the 2018 Artificial Intelligence and Cloud Computing Conference. <https://doi.org/10.1145/3299819.3299841>

32. Chang, Y., Government Information Quarterly (2018), <https://doi.org/10.1016/j.giq.2018.04.002>
33. El Haurani L., Abou El Kalam A., Ait Ouahman A. (2019) Knowledge Based Access Control a Model for Security and Privacy in the Big Data. In: Ben Ahmed M., Boudhir A., Younes A. (eds) Innovations in Smart Cities Applications Edition 2. SCA 2018. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham. https://doi.org/10.1007/978-3-030-11196-0_60
34. N. Miloslavskaya, A. Nikiforov and V. Budzko, Standardization of Ensuring Information Security for Big Data Technologies, 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), 2018, pp. 56-63, <https://doi.org/10.1109/W-FiCloud.2018.00015>
35. Reza Saneei Moghadam, Ricardo Colomo-Palacios. Information security governance in big data environments: A systematic mapping, Procedia Computer Science, Volume 138, 2018, Pages 401-408, <https://doi.org/10.1016/j.procs.2018.10.057>
36. Khairulliza Ahmad Salleh, Lech Janczewski, Security Considerations in Big Data Solutions Adoption: Lessons from a Case Study on a Banking Institution, Procedia Computer Science, Volume 164, 2019, Pages 168-176, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.12.169>
37. Shetty M.M., Manjaiah D.H., Hemdan E.ED. (2019) Policy-Based Access Control Scheme for Securing Hadoop Ecosystem. In: Balas V., Sharma N., Chakrabarti A. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing, vol 839. Springer, Singapore. https://doi.org/10.1007/978-981-13-1274-8_13
38. M. Tang, M. Alazab and Y. Luo, Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies, in IEEE Transactions on Big Data, vol. 5, no. 3, pp. 317-329, 1 Sept. 2019, <https://doi.org/10.1109/TBDATA.2017.2723570>
39. F. Xhafa et al. (eds.), Research into Information Security Strategy. Practices for Commercial Banks in Taiwan Recent Developments in Intelligent Systems and Interactive Applications, Advances in Intelligent Systems and Computing 541, https://doi.org/10.1007/978-3-319-49568-2_25
40. A. Yarali, R. Joyce and B. Dixon, Ethics of Big Data: Privacy, Security and Trust, 2020 Wireless Telecommunications Symposium (WTS), 2020, pp. 1-7, <https://doi.org/10.1109/WTS48268.2020.9198734>
41. Sahel Alouneh, Ismail Hababeh, Tamer Alajrami. Toward big data analysis to improve enterprise information security. MEDES '18: Proceedings of the 10th International Conference on Management of Digital EcoSystems September 2018, Pages 106–109, <https://doi.org/10.1145/3281375.3281393>
42. Macklin, Thomas ; Mathews, Joseph. Big data, little security: Addressing security issues in your platform. Proceedings of the SPIE, Volume 10207, id. 102070G 10 pp. (2017). <https://doi.org/10.1117/12.2268002>
43. Junjun Guo, Le Wang, Learning to upgrade internet information security and protection strategy in big data era, Computer Communications, Volume 160, 2020, Pages 150-157, <https://doi.org/10.1016/j.comcom.2020.05.043>
44. Maanak Gupta, Farhan Patwa, Ravi Sandhu . An Attribute-Based Access Control Model for Secure Big Data Processing in Hadoop Ecosystem. the Third ACM Workshop. <https://doi.org/10.1145/3180457.3180463>
45. Colombo, P., Ferrari, E. Access control technologies for Big Data management systems: literature review and future trends. Cybersecur 2, 3 (2019). <https://doi.org/10.1186/s42400-018-0020-9>
46. N. Lee and B. Wu, Privacy Protection Technology and Access Control Mechanism for Medical Big Data, 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017, pp. 424-429, <https://doi.org/10.1109/IIAI-AAI.2017.34>
47. D. Mondek, R. B. Blažek and T. Zahradnický, "Security Analytics in the Big Data Era," 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2017, pp. 605-606, <https://doi.org/10.1109/QRS-C.2017.136>
48. Meisuchi Naisuty, Achmad Nizar Hidayanto, Nabila Clydea Harahap, Ahmad Rosyiq, Agus Suhanto and George Michael Samuel Hartono. Data protection on hadoop distributed file system by using encryption algorithms: a systematic literature review. 2020 J. Phys.: Conf. Ser. 1444 012012. <https://doi.org/10.1088/1742-6596/1444/1/012012>
49. Salas, J., Domingo-Ferrer, J. Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges. Math.Comput.Sci. 12, 263–274 (2018). <https://doi.org/10.1007/s11786-018-0344-6>

51. Hai Tao, Md Zakirul Alam Bhuiyan, Md Arafatur Rahman, Guojun Wang, Tian Wang, Md. Manjur Ahmed, Jing Li, Economic perspective analysis of protecting big data security and privacy, *Future Generation Computer Systems*, Volume 98, 2019, Pages 660-671, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2019.03.042>
52. Linqi Zhou, Weihong Gu, Cheng Huang, Aijun Huang, and Yongbin Bai. Research on information security in big data era. *AIP Conference Proceedings* 1967, 020020 (2018). <https://doi.org/10.1063/1.5038992>
53. Poltavtseva, M.A., Zegzhda, D.P. & Kalinin, M.O. Big Data Management System Security Threat Model. *Aut. Control Comp. Sci.* 53, 903–913 (2019). <https://doi.org/10.3103/S0146411619080261>
54. GOST R 53114-2008. Zashhita informacii. Obespechenie informacionnoj bezopasnosti v organizacii. Osnovny`e terminy` i opredeleniya. Utverzhen i vveden v dejstvie Prikazom Federal`nogo agentstva po texnicheskomu regulirovaniyu i metrologii ot 18 dekabrya 2008 g. N 532-st

Цифровое сельское хозяйство**АНАЛИЗ ЦИФРОВЫХ ТЕХНОЛОГИЙ В АГРОПРОДОВОЛЬСТВЕННОЙ
СФЕРЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ОБРАБОТКИ БОЛЬШИХ
ДАННЫХ**

Статья рекомендована к публикации членом редакционного совета А.Н. Райковым 29.09.2021.

Отмахова Юлия Сергеевна

Кандидат экономических наук

*Центральный экономико-математический институт РАН, лаборатория компьютерного моделирования социально-экономических процессов, ведущий научный сотрудник
Москва, Россия
otmakhovajs@yandex.ru*

Девяткин Дмитрий Алексеевич

*Федеральный исследовательский центр «Информатика и управление» РАН, отдел интеллектуальных технологий и систем, научный сотрудник
Москва, Россия
devyatkin@isa.ru*

Усенко Наталья Ивановна

Кандидат экономических наук

*ООО «Технологии системного анализа», ведущий научный сотрудник
Москва, Россия
n.i.usenko@yandex.ru*

Аннотация

Цифровые технологии становятся все более востребованными различными отраслями экономики в условиях пандемии COVID-19. Анализ практики использования подобных технологий выявил необходимость в теоретическом осмыслении возможностей формирования новой цифровой модели и исследовании потребностей агропродовольственного комплекса в цифровых технологиях. В работе представлен подход к анализу цифровых технологий в агропродовольственной сфере на примере робототехнических систем. В результате исследования сформирована коллекция полнотекстовых документов по робототехнике, которая позволила выявить динамику и структуру патентов и публикаций по данной тематике. Результаты исследования представляют основу для развития методологии интеллектуального анализа больших данных применительно к технологиям цифровизации. Результаты могут быть использованы при определении инструментов государственного управления в процессе принятия решений по ключевым индикаторам перспективных направлений и рынков, а также при определении наиболее перспективных направлений научно-технологического развития агропромышленного комплекса.

Ключевые слова

Цифровые технологии; анализ больших данных; патентный ландшафт; пищевая промышленность; сельское хозяйство

Введение

Современный мир стремительно меняется, изменяются условия функционирования экономики, каждого социального института, каждого человека. Одним из таких факторов глубоких

© Отмахова Ю.С., Девяткин Д.А., Усенко Н.И. 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_334

изменений является цифровизация всех сторон развития общества, меняются потребности отраслей экономики в цифровых платформах и технологиях. Передовые технологические направления, такие как искусственный интеллект, робототехника, блокчейн, технологии виртуальной и дополненной реальности, стали важнейшим катализатором нового этапа цифровой трансформации [4, 21]. Изменение современного мира под влиянием цифровизации имеет глобальный характер, а происходящие цифровые трансформации становятся факторами глобальной конкуренции. В условиях пандемии COVID-19 (коронавирус нового типа) ландшафт рынка труда претерпел существенные изменения, происходит формирование новых профессий под влиянием цифровизации. В статье [1] показано влияние развития цифровых информационных технологий на рынок труда и занятость населения. Агропродовольственный комплекс является важнейшим сектором мировой экономики, при этом применяющиеся практики цифровизации сельского хозяйства и продовольственного комплекса опережают методологические и методические экономические исследования в этой области [6]. В статье [7] показано, что проблемы разнонаправленных интересов участников продовольственного рынка приводят к различным трансформациям на рынках социально-значимой пищевой продукции. В статье [8] изложены подходы к вопросам методологии формирования современных цифровых платформенных моделей бизнеса, в том числе инструментария для измерения и оценки готовности участников агропродовольственного рынка к цифровой трансформации. В статье [9] представлена сегментация потребностей отраслей в сквозных цифровых технологиях и субтехнологиях, что особенно актуально в рамках агропродовольственного рынка в условиях создающихся цифровых платформ агропромышленного комплекса в России.

Теоретические аспекты формирования новой цифровой модели по данному сектору экономики и анализ потребностей в цифровых технологиях в публикациях представлены весьма слабо. Динамичное развитие робототехнической сферы актуализирует необходимость исследования данной технологической области. В этих условиях исследование важнейших направлений цифровизации на мировом уровне актуализирует необходимость использования новых подходов к анализу данных, в том числе современных методов и инструментов интеллектуального анализа. Методы построения патентных и научных ландшафтов являются важным инструментом для оценки технологических тенденций и анализа конкурентного ландшафта в различных отраслях экономики. Анализ патентного и публикационного ландшафтов цифровых технологий в агропродовольственной сфере позволит экспертам опираться на конкретные данные по патентной активности в мировом масштабе и тем самым дает возможность более обоснованно формулировать тренды развития цифровых технологий в России.

Целью настоящего исследования является выполнение анализа современных цифровых технологий в агропродовольственной сфере с использованием методов интеллектуального анализа больших данных. В рамках исследования были поставлены следующие задачи: разработать методику формирования коллекции полнотекстовых документов в виде патентов и публикаций; разработать подход к анализу цифровых технологий с использованием методов интеллектуального анализа больших данных; провести анализ патентной активности и выявить основных правообладателей патентов в категории «Робототехнические системы в сельском хозяйстве, производстве и поставках продовольствия» в базе патентов USPTO (United States Patent and Trademark Office); провести анализ публикационной активности и структуру публикаций по авторам исследований в рамках базы Scopus.

Новизна работы заключается в новом подходе к оценке цифровых технологий, который позволяет проводить интеграцию методов экономического анализа и методов анализа больших данных по патентной и публикационной активности.

В данной работе в качестве объекта исследования было выбрано такое важнейшее направление цифровизации как робототехнические системы в сельском хозяйстве, производстве и поставках продовольствия.

Выбор данного направления объясняется возрастающей практической значимостью для решения задач повышения производительности и эффективности производства, а также повышения качества выпускаемой продукции. Промышленные роботы имеют высокий потенциал применений для пищевой промышленности, а благодаря машинному зрению, технологиям искусственного интеллекта и внедрению новых материалов современные промышленные роботы характеризуются высокой степенью гибкости по отношению к широкому спектру задач. На пищевых предприятиях задачи маркировки, упаковки, складирования могут быть эффективно

выполнены с помощью роботизированных систем для получения эффективных результатов. Для сельского хозяйства разработаны и разрабатываются роботы для высаживания цветов, для прореживания растений, для сбора урожая и для обрезки растений, для погрузки-разгрузки, перемещения контейнеров, для уборки кукурузы, пшеницы и соевых бобов, для внесения удобрений, дроны для точного земледелия и другое. Особую практическую значимость исследования придает его непосредственная связь с целями и задачами Национальной программы «Цифровая экономика Российской Федерации», которая предусматривает возможности применения цифровых технологий в пищевой индустрии и сельском хозяйстве для реализации принципов «Индустрия 4.0».

1 Методы и инструментарий исследования

В данной работе используются подходы к построению научных и патентных ландшафтов (картированию науки) с использованием методов информационного поиска и интеллектуального анализа текстов публикаций. Научные ландшафты демонстрируют тенденции развития в выбранной области знаний [2]. Картирование науки выполняется для определенной предметной области, а сам процесс анализа полнотекстовых публикаций, найденных с помощью информационно-аналитических систем (Google Patents, PatSearch, Exactus Patent, Scopus, Web of Science и др.) проводится с использованием специализированного программного обеспечения [14, 21], позволяющего извлекать из научных текстов термины, именованные сущности, проводить анализ графов цитирований. Результаты картирования представлены в работах [11, 12, 13, 17, 18, 20].

Создание патентного ландшафта включает статистический и интеллектуальный анализ патентных документов, который позволяет получить объективные данные о современном состоянии областей техники и перспектив их развития для целостного понимания технических и конкурентных тенденций на рынке или в конкретном техническом пространстве [3, 5]. В работе [15] представлены результаты технологий проектирования патентных ландшафтов для выявления ключевых центров компетенции для оценки существующего научно-технического потенциала для расширения экспорта. Следует отметить работы, которые при построении научного и патентного ландшафтов используют методы машинного обучения и кластеризации для поиска похожих проектов по разным научным областям [10, 14].

Научно-технический ландшафт комбинирует оба перечисленных типа ландшафтов. Он позволяет выявить траектории развития технологий, отследить основные пути создания новых технологий на основе полученных ранее научных результатов. Основная проблема, связанная с построением подобных ландшафтов, состоит в необходимости интеграции баз разнородных научно-технических документов, что на практике затруднено из-за невозможности создания единой (сквозной) системы классификации научно-технических документов, а также применения различного, несвязанного программного обеспечения для работы с этими базами.

В работе предложена авторская методика, позволяющая интегрировать разнородные базы научно-технических документов и выполнить их совместный анализ. Методика, таким образом, дает возможность сформировать научно-технический ландшафт в заданной сфере деятельности. В основе этой методики лежит совместное использование поисково-аналитической системы ИАС «Приоритеты», с помощью которой выполняется анализ базы патентов США (USPTO) и наукометрической базы данных Scopus для сбора и анализа документов по тематике «робототехнические системы в агропродовольственной сфере». Предложенная методика анализа патентной и публикационной активности была использована для оценки уровня и масштаба проводимых международных исследований и патентования по робототехнике. Анализ научно-технической деятельности на международном уровне позволяет снизить влияние изменений учетной политики в научных организациях отдельных стран на результаты исследования. Исследование включало выполнение следующих этапов (рисунки 1):

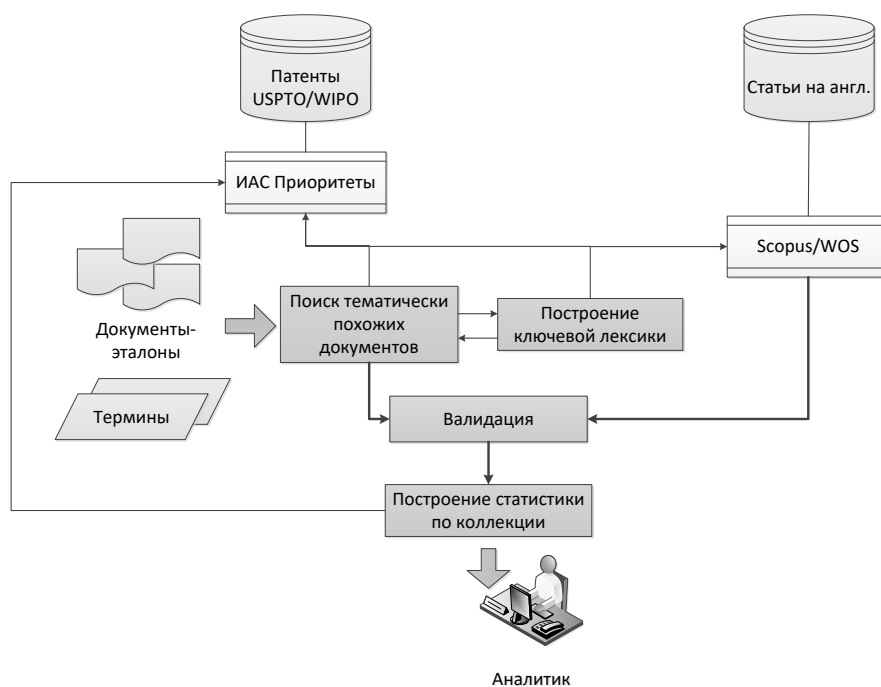


Рисунок 1. Схема формирования агрегированной коллекции разнородных научно-технических документов (патентов и научных публикаций) для анализа технологий в заданной сфере деятельности

1. Формирование набора документов-эталонов (seeds) по каждой технологии. Набор формируется с помощью системы ИАС «Приоритеты» с привлечением экспертов в анализируемой предметной области. Для автоматизации этого процесса используется полнотекстовый семантический поиск научно-технических документов по запросу.
2. Построение ключевой лексики для технологий и выбор базы данных для наукометрического и патентного анализа. Для построения лексики используется метод, основанный на оценке характеристики тематической значимости лексико-синтаксических элементов (слов и словосочетаний) текстов документов-эталонов [19].
3. Формирование коллекций научных публикаций. Для построения коллекций используются поисковые запросы, составленные на основе ключевой лексики, полученной на шаге 2. Далее эти запросы используются для поиска публикаций в системе Scopus.
4. Формирование коллекций полнотекстовых документов (патентов). В качестве основы для формирования коллекций используются документы-эталоны, найденные на шаге 1. Для формирования коллекции используются методы поиска тематически близких документов, реализованные в системе ИАС «Приоритеты» [22]. Процесс расширения коллекции производится итеративно. На каждом шаге аналитик выполняет валидацию результатов поиска и отфильтровывает нерелевантные документы. Процесс продолжается до тех пор, пока по итогам валидации удастся добавить в коллекцию хотя бы один релевантный документ. В качестве временного интервала для поиска публикаций и патентов был выбран промежуток за последние десять лет (с 2010 по 2021 гг.).
5. Детальный анализ патентной и публикационной активности по выбранной цифровой технологии. Этот этап включает построение агрегированной статистики по метаданным документов из сформированных коллекций, включающей годы публикации, основных владельцев патентов, авторов статей и их аффилиации. Полученная информация в совокупности с экспертными данными, данными о финансировании исследований и разработок в анализируемых областях может быть использована для оценки уровня готовности технологий и производства (УГТ, УПП) по ГОСТ Р 58048–2017.

2 Результаты исследования

В рамках исследования была сформирована коллекция полнотекстовых документов по робототехнике, включая 1585 патентов USPTO и 704 публикации базы Scopus. В работе

представлен авторский подход к анализу цифровых технологий и в результате проведенного исследования была сформирована ключевая лексика по тематике робототехнические системы в сельском хозяйстве, производстве и поставках продовольствия. На нижеприведенном рисунке 2 представлено полученное облако тегов, которое позволяет выделить акцент на важных словах, которые представлены в научных публикациях и патентах.



Рисунок 2. Облако ключевых слов и словосочетаний

В рамках исследования показано, что за последние десять лет происходит постоянный рост патентной активности в области создания и внедрения робототехнических систем для использования в агропродовольственном комплексе (рисунок 3). Следует отметить, что получение патентов финансово и организационно затратная процедура, требующая значительных усилий и времени, чем режим охраны изобретений в виде коммерческой тайны (ноу-хау).

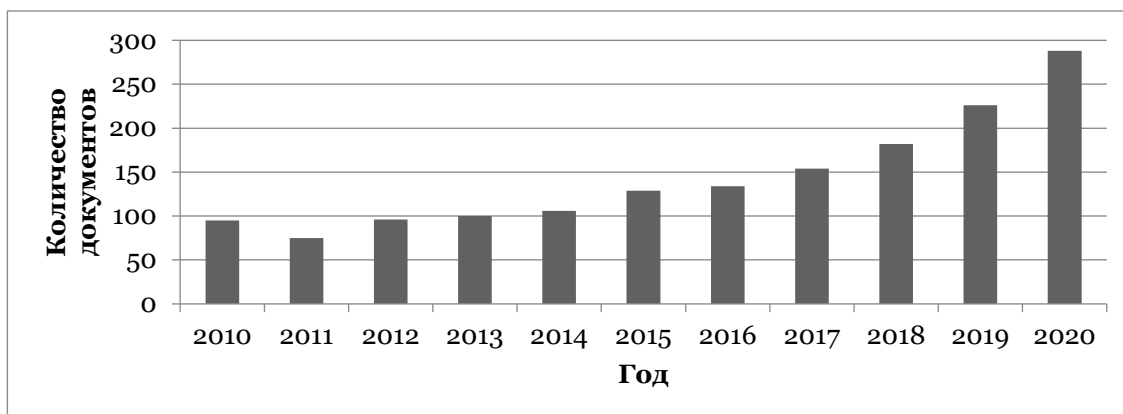


Рисунок 3. Динамика патентной активности в период 2010–2020 гг.

В работе был проведен анализ 1585 патентов базы USPTO, которые были сосредоточены в 196 компаниях-патентообладателях. Анализ государственной принадлежности патентообладателей показал, что около 72% патентов принадлежит американским компаниям, далее со значительным отрывом следуют Япония и Германия с долей патентов 12% и 9% соответственно.

В данной работе была решена задача по выявлению основных правообладателей патентов в категории «Робототехнические системы в сельском хозяйстве, производстве и поставках продовольствия» в базе патентов USPTO. По результатам полученного патентного ландшафта в данном исследовании графически представлен рэнкинг основных патентообладателей в области робототехнических систем в агропродовольственной сфере (рисунок 4) по состоянию на 2020 год. Анализ полученных результатов показал, что безусловным лидером по количеству патентов является известная американская компания «DOW Agro Sciences», занимающая лидирующие позиции в мире по изготовлению и реализации высококачественных средств защиты растений. Компания имеет научно-производственный потенциал, владеет 44 патентами USPTO по робототехническим системам для использования в агропродовольственном комплексе.

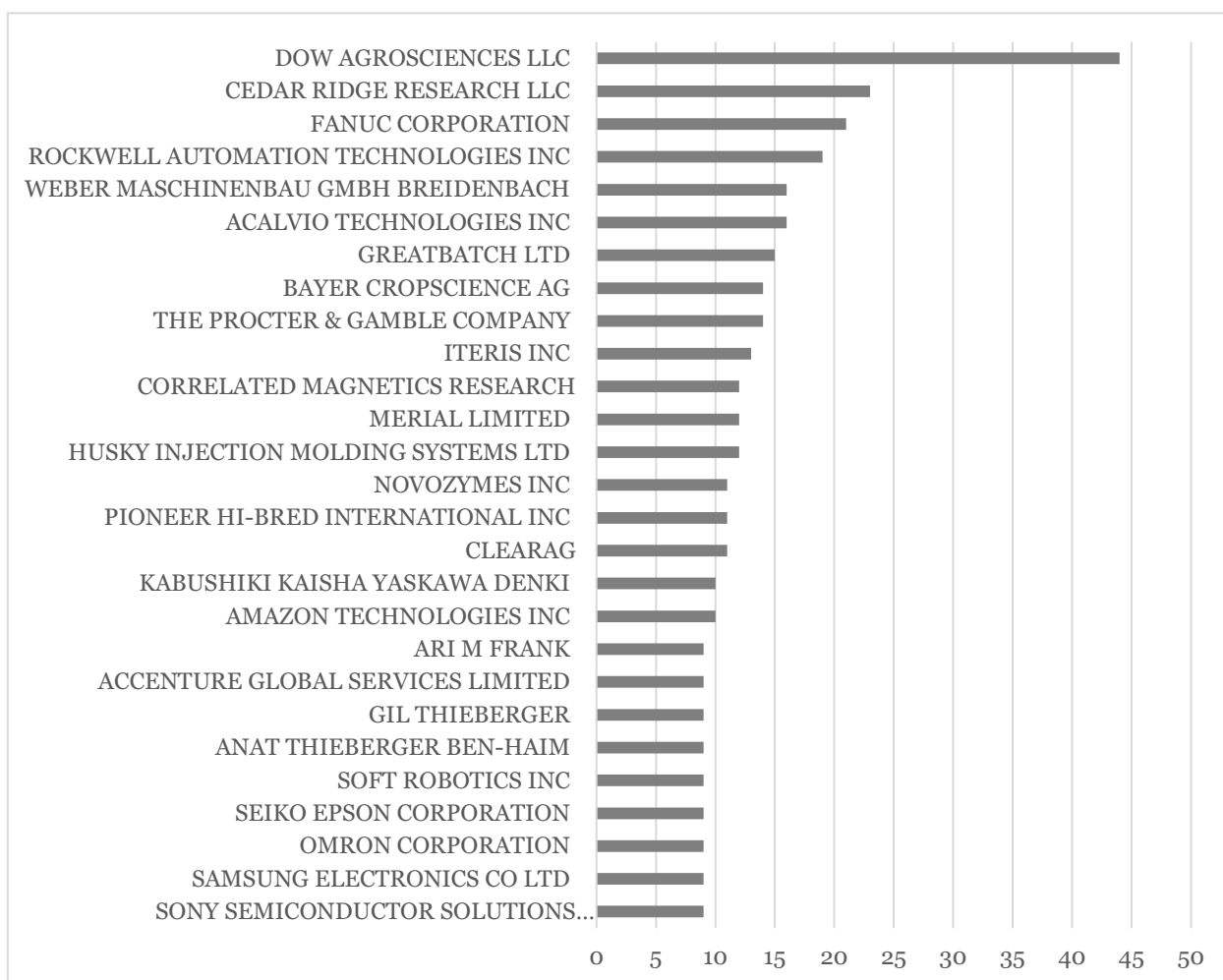


Рисунок 4. Основные патентообладатели (USPTO)

Второе и третье места разделяют такие компании как Cedar Ridge Research и Fanuc Corporation. Американская глобальная консалтинговая компания Cedar Ridge Research предлагает услуги в области интеллектуальной собственности и владеет 23 патентами USPTO. Японская технологическая компания FANUC является мировым лидером в автоматизации и роботизации промышленных предприятий, владеет 21 патентами. FANUC производит пищевых роботов, которая позволяет автоматизировать широкий ряд операций в пищевой промышленности и производстве напитков. Четвертое место в ренкинге занимает американская компания Rock Well Automation Technologies, которая имеет 19 патентов по данной тематике и является мировым лидером в области промышленной автоматизации и информационных продуктов в таких сегментах пищевого производства как молочные продукты, хлебопечение, производство напитков, пивоварен и полуфабрикатов.

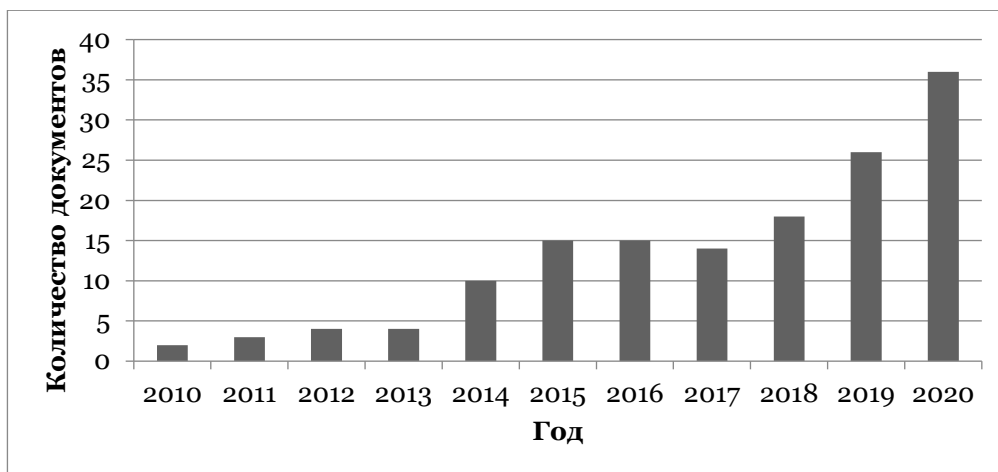


Рисунок 5. Динамика публикационной активности

В рамках решения задачи по оценке международной публикационной активности по направлению робототехнические системы в сельском хозяйстве, производстве и поставках продовольствия, которая отражает уровень развития данного направления науки, был проведен анализ количества публикаций в базе Scopus (рисунок 5). Анализ публикаций по выбранной тематике включал 704 работы, который продемонстрировал неуклонной рост количества статей с каждым годом, что свидетельствует об активной развивающейся тематике исследования по применению робототехнических систем в сельском хозяйстве и пищевой индустрии. Анализ структуры публикаций показал, что по данной тематике большинство авторов за период 2010–2020 гг. имеют по одной публикации, что свидетельствует становлении нового научного направления.

Заключение

Происходящие цифровые трансформации становятся факторами глобальной конкуренции, поэтому понимание современного состояния науки и технологий в области цифровизации агропродовольственного комплекса с использованием методов научного и патентного ландшафта важно для формирования приоритетов в определении государственных заданий и тематики грантовых программ для научно-исследовательских организаций и высших учебных заведений.

В работе показано, что на мировом рынке среди патентообладателей, лидирующих по количеству патентов по робототехнике в агропродовольственной сфере, можно выделить мощные международные диверсифицированные компании, имеющие развитую сеть своих подразделений по всему миру, а патенты являются лишь составной частью продуктов корпораций. Результаты данного исследования позволили оценить уровень и масштаб публикационной и патентной активности международных исследований по робототехнике как перспективного направления цифровизации, способствующего формированию новой реальности в агропродовольственной сфере. Данные результаты могут быть учтены при формировании государственных программ по агропродовольственной политике с учетом глобальных технологических трендов в рамках поддержки прорывных технологий с позиций национальной продовольственной безопасности.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 18-29-03086. Исследование выполнено также при частичной финансовой поддержке проекта «Средства интеллектуального анализа больших массивов текстов» в рамках программы Центров компетенций Национальной технологической инициативы на базе Московского государственного университета имени М.В.Ломоносова (соглашение о финансовой поддержке проектов НТИ № 13/1251/2018 от 11.12.2018).

Литература

1. Абдулхаирова Э. М. Последствия цифровизации экономики для занятости населения // Наука Красноярья. 2020. Т. 9. № 4. С. 7–30. DOI 10.12731/2070-7568-2020-4-7-30.
2. Анализ научного и патентного ландшафтов в сфере современных технологий глубокой переработки зерна / Отмахова Ю. С., Крескин А. Д., Девяткин Д. А., Тихомиров И. А. // Инновации. 2020. № 2(256). С. 89–96. DOI 10.26310/2071-3010.2020.256.2.012.
3. Анализ научного и патентного ландшафтов современных радиационных технологий облучения пищевых продуктов и сырья / Отмахова Ю. С., Девяткин Д. А., Крескин А. Д., Усенко Н. И. // Информационное общество. 2020. № 1. С. 57–70.
4. Ершова Т. В., Райков А. Н., Хохлов Ю. Е. Система мониторинга потребностей отраслей экономики в цифровых платформах и технологиях // Информационное общество. 2020. № 2. С. 2–19.
5. Кузьминов И. Ф., Логинова И. В., Лобанова П. А. Перспективы использования технологий анализа больших данных для стратегической аналитики агропромышленного комплекса // Сахарная свекла. 2018. № 9. С. 9–15. DOI 10.25802/SB.2018.56.64.002.
6. Отмахова, Ю. С., Асавасанти С. Возможности цифровой трансформации в пищевой индустрии // Мир экономики и управления. 2020. Т. 20. № 1. С. 116–127. DOI 10.25205/2542-0429-2020-20-1-116-127.
7. Отмахова Ю. С., Ибрагимов Н.М. Агент-ориентированное моделирование поведения потребителей на рынке молочной продукции // Наука Красноярья. 2020. Т. 9. № 4. С. 281–295. DOI 10.12731/2070-7568-2020-4-281-295.
8. Отмахова Ю. С., Усенко Н. И. Цифровизация и новые подходы к управлению агропродовольственным комплексом // Пищевая промышленность. 2019. № 12. С. 35–38. DOI 10.24411/0235-2486-2019-10192.
9. Raikov, A.N., Ermakov, A.N., and Merkulov, A.A., Assessments of the Economic Sectors Needs in Digital Technologies, Lobachevskii Journal of Mathematics, 2019, Vol. 40, No. 11, pp. 1837–1847. Pleiades Publishing, Ltd. Doi: <https://doi.org/10.1134/S1995080219110246>.
10. Abood A., Feltenberger D. Automated patent landscaping. *Artificial Intelligence and Law*, 2018, vol. 26, 2, p. 103-125. DOI 10.1007/s10506-018-9222-4.
11. Agarwal A., Baskaran S., Selvam M. K. P., Finelli R., Barbarosie C., Robert K. A, Henkel R. Scientific landscape of oxidative stress in male reproductive research: A scientometric study. *Free Radical Biology and Medicine*, 2020, vol. 156, p. 36-44. DOI 10.1016/j.freeradbiomed.2020.05.008.
12. Bañales-Seguel C., Riquelme Maulén W., Álvez, A., Habit, E. Scientific Landscape Related to Mapuche Indigenous Peoples and Wallmapu Territory. *Sustainability*, 2020, vol. 12, 19, p. 7895. DOI 10.3390/su12197895.
13. Castor K., Mota F. B., Da Silva R. M., Cabral B. P., Maciel E. L., De Almeida I. N., Kritski A. Mapping the tuberculosis scientific landscape among BRICS countries: a bibliometric and network analysis. *Memórias do Instituto Oswaldo Cruz*, 2020, p. 115. DOI 10.1590/0074-02760190342.
14. Kawamura T., Watanabe K., Matsumoto N., Egami S., Jibu M. Science Graph for characterizing the recent scientific landscape using Paragraph Vectors. In *Proceedings of the Knowledge Capture Conference*, 2017, p. 1–8. DOI 10.1145/3148011.3148018.
15. Matos F. L., Ross S. W., Huvenne V. A. I., Davies J. S., Cunha M. R. Canyons pride and prejudice: Exploring the submarine canyon research landscape, a history of geographic and thematic bias. *Progress in Oceanography*, 2018, vol. 169, p. 6–19. DOI 10.1016/j.pocean.2018.04.010.
16. Otmakhova Yu. S., Usenko N. I., Devyatkin D. A. [et al.] Agri-food export: challenges and structural changes. Novosibirsk: Novosibirsk state research university, 2021, 174 p., ISBN 9785443709802.
17. Popkirov S., Jungilligens J., Schlegel U., Wellmer J. Research on dissociative seizures: a bibliometric analysis and visualization of the scientific landscape. *Epilepsy and Behavior*, 2018, vol. 83, p. 162–167. DOI 10.1016/j.yebeh.2018.03.041.
18. Sheikhnejad Y., Yigitcanlar T. Scientific landscape of sustainable urban and rural areas research: A systematic scientometric analysis. *Sustainability*, 2020, vol.12 (4), p. 1293. DOI 10.3390/su12041293.
19. Shvets A. et al. Detection of current research directions based on full-text clustering. 2015 *Science and Information Conference (SAI)*. – IEEE, 2015 – С. 483-488.

20. Yeung A. W. K., Heinrich M., Kijjoa A., Tzvetkov N. T., Atanasov A. G. The ethnopharmacological literature: An analysis of the scientific landscape. *Journal of ethnopharmacology*, 2020, vol. 250, p. 112414. DOI 10.1016/j.jep.2019.112414.
21. Zhang G., Morris E., Allaire D., and McAdams D.A. Research Opportunities and Challenges in Engineering System Evolution. *Journal of Mechanical Design*, vol. 142, № 8, 2020.
22. Zubarev D.V., Sochenkov I.V. Cross-lingual similar document retrieval methods. *Proceedings of the Institute for System Programming of the RAS*. 2019. T. 31. № 5. С. 127 - 136.

ANALYSIS OF DIGITAL TECHNOLOGIES IN THE AGRI-FOOD SECTOR BASED ON BIG DATA ANALYSIS METHODS

Otmakhova, Yulia Sergeevna

Candidate of economic sciences

Central Economic and Mathematics Institute of the Russian Academy of Sciences, Laboratory of computer modeling of socio-economic processes, leading researcher

Moscow, Russia

otmakhovajs@yandex.ru

Devyatkin, Dmitry Alexeevich

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Department of intelligent technologies and systems, researcher

Moscow, Russia

devyatkin@isa.ru

Usenko, Natalia Ivanovna

Candidate of economic sciences

Technologies for Systems Analysis LLC, leading researcher

Moscow, Russia

n.i.usenko@yandex.ru

Abstract

Digital technologies are becoming more and more in demand by various sectors of the economy because of the COVID-19 pandemic. The study of the digital technologies shows a need for a theoretical studying of forming a new digital model and analyzing the requirements for those technologies in the agriculture. The paper presents an approach to analyzing digital technologies in the agriculture, including robotic systems. As a result, we have formed a collection of full-text documents on robotics. The analysis of that collection reveals the dynamics and structure of international patenting and publications on that topic. The obtained results form the basis for developing the methodology of big data mining to such critical digitalization technologies as robotic systems.

Keywords

Digital technologies; big data analysis; patent landscape; food industry; agriculture

References

1. Abdulkhairova E. M. Posledstviya tsifrovizatsii ekonomiki dlya zanyatosti naseleniya [Consequences of the digitalization of the economy for the employment of the population] // Krasnoyarsk Science. 2020. Vol. 9. no 4. pp. 7–30. DOI 10.12731/2070–7568-2020-4-7-30.
2. Otmakhova Yu. S., Kreskin A. D., Devyatkin D. A., Tikhomirov I. A. Analiz nauchnogo i patentnogo landshaftov v sfere sovremennykh tekhnologiy glubokoy pererabotki zerna [Analysis of scientific and patent landscapes in the field of modern technologies for deep processing of grain] // Innovations. 2020. no 2(256). pp. 89–96. DOI 10.26310/2071–3010.2020.256.2.012.
3. Otmakhova Yu. S., Devyatkin D. A., Kreskin A. D., Usenko N. I. Analiz nauchnogo i patentnogo landshaftov sovremennykh radiatsionnykh tekhnologiy oblucheniya pishchevykh produktov i syr'ya [Analysis of scientific and patent landscapes of modern radiation technologies for irradiation of food products and raw materials] // Information Society. 2020. no 1. pp. 57–70.
4. Ershova T. V., Raykov A. N., Khokhlov Yu. E. Sistema monitoringa potrebnostey otrasley ekonomiki v tsifrovyykh platformakh i tekhnologiyakh [System for monitoring the needs of economic sectors in digital platforms and technologies] // Information Society. 2020. no 2. pp. 2–19.
5. Kuz'minov I. F., Loginova I. V., Lobanova P. A. Perspektivy ispol'zovaniya tekhnologiy analiza bol'shikh dannykh dlya strategicheskoy analitiki agropromyshlennogo kompleksa [Prospects for the use of big data analysis technologies for strategic analytics of the agro-industrial complex] // Sugar Beet. 2018. no 9. pp. 9–15. DOI 10.25802/SB.2018.56.64.002.

6. Otmakhova, Yu. S., Asavasanti S. *Vozmozhnosti tsifrovoy transformatsii v pishchevoy industrii* [Opportunities for digital transformation in the food industry] // *World of Economics and Management*. 2020. Vol. 20. no 1. pp. 116–127. DOI 10.25205/2542-0429-2020-20-1-116-127.
7. Otmakhova Yu. S., Ibragimov N.M. *Agent-orientirovannoe modelirovanie povedeniya potrebiteley na rynke molochnoy produktsii* [Agent-based modeling of consumer behavior in the dairy market] // *Krasnoyarsk Science*. 2020. Vol. 9. no 4. pp. 281–295. DOI 10.12731/2070-7568-2020-4-281-295.
8. Otmakhova Yu. S., Usenko N. I. *Tsifrovizatsiya i novye podkhody k upravleniyu agroproduktstvennym kompleksom* [Digitalization and new approaches to managing the agri-food complex] // *Food Industry*. 2019. no 12. pp. 35–38. DOI 10.24411/0235-2486-2019-10192.
9. Raikov, A.N., Ermakov, A.N., and Merkulov, A.A., *Assessments of the Economic Sectors Needs in Digital Technologies*, *Lobachevskii Journal of Mathematics*, 2019, Vol. 40, No. 11, pp. 1837–1847. Pleiades Publishing, Ltd. Doi: <https://doi.org/10.1134/S1995080219110246>.
10. Abood A., Feltenberger D. *Automated patent landscaping*. *Artificial Intelligence and Law*, 2018, vol. 26, 2, p. 103-125. DOI 10.1007/s10506-018-9222-4.
11. Agarwal A., Baskaran S., Selvam M. K. P., Finelli R., Barbarosie C., Robert K. A, Henkel R. *Scientific landscape of oxidative stress in male reproductive research: A scientometric study*. *Free Radical Biology and Medicine*, 2020, vol. 156, p. 36-44. DOI 10.1016/j.freeradbiomed.2020.05.008.
12. Bañales-Seguel C., Riquelme Maulén W., Álvez, A., Habit, E. *Scientific Landscape Related to Mapuche Indigenous Peoples and Wallmapu Territory*. *Sustainability*, 2020, vol. 12, 19, p. 7895. DOI 10.3390/su12197895.
13. Castor K., Mota F. B., Da Silva R. M., Cabral B. P., Maciel E. L., De Almeida I. N., Kritski A. *Mapping the tuberculosis scientific landscape among BRICS countries: a bibliometric and network analysis*. *Memórias do Instituto Oswaldo Cruz*, 2020, p. 115. DOI 10.1590/0074-02760190342.
14. Kawamura T., Watanabe K., Matsumoto N., Egami S., Jibu M. *Science Graph for characterizing the recent scientific landscape using Paragraph Vectors*. In *Proceedings of the Knowledge Capture Conference, 2017*, p. 1–8. DOI 10.1145/3148011.3148018.
15. Matos F. L., Ross S. W., Huvenne V. A. I., Davies J. S., Cunha M. R. *Canyons pride and prejudice: Exploring the submarine canyon research landscape, a history of geographic and thematic bias*. *Progress in Oceanography*, 2018, vol. 169, p. 6–19. DOI 10.1016/j.pocean.2018.04.010.
16. Otmakhova Yu. S., Usenko N. I., Devyatkin D. A. [et al.] *Agri-food export: challenges and structural changes*. Novosibirsk: Novosibirsk state research university, 2021, 174 p., ISBN 9785443709802.
17. Popkirov S., Jungilligens J., Schlegel U., Wellmer J. *Research on dissociative seizures: a bibliometric analysis and visualization of the scientific landscape*. *Epilepsy and Behavior*, 2018, vol. 83, p. 162–167. DOI 10.1016/j.yebeh.2018.03.041.
18. Sheikhnejad Y., Yigitcanlar T. *Scientific landscape of sustainable urban and rural areas research: A systematic scientometric analysis*. *Sustainability*, 2020, vol.12 (4), p. 1293. DOI 10.3390/su12041293.
19. Shvets A. et al. *Detection of current research directions based on full-text clustering*. 2015 *Science and Information Conference (SAI)*. – IEEE, 2015 – pp. 483-488
20. Yeung A. W. K., Heinrich M., Kijjoa A., Tzvetkov N. T., Atanasov A. G. *The ethnopharmacological literature: An analysis of the scientific landscape*. *Journal of ethnopharmacology*, 2020, vol. 250, p. 112414. DOI 10.1016/j.jep.2019.112414.
21. Zhang G., Morris E., Allaire D., and McAdams D.A. *Research Opportunities and Challenges in Engineering System Evolution*. *Journal of Mechanical Design*, vol. 142, № 8, 2020.
22. Zubarev D.V., Sochenkov I.V. *Cross-lingual similar document retrieval methods*. *Proceedings of the Institute for System Programming of the RAS*. 2019. T. 31. № 5. pp. 127–136.