

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 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 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.

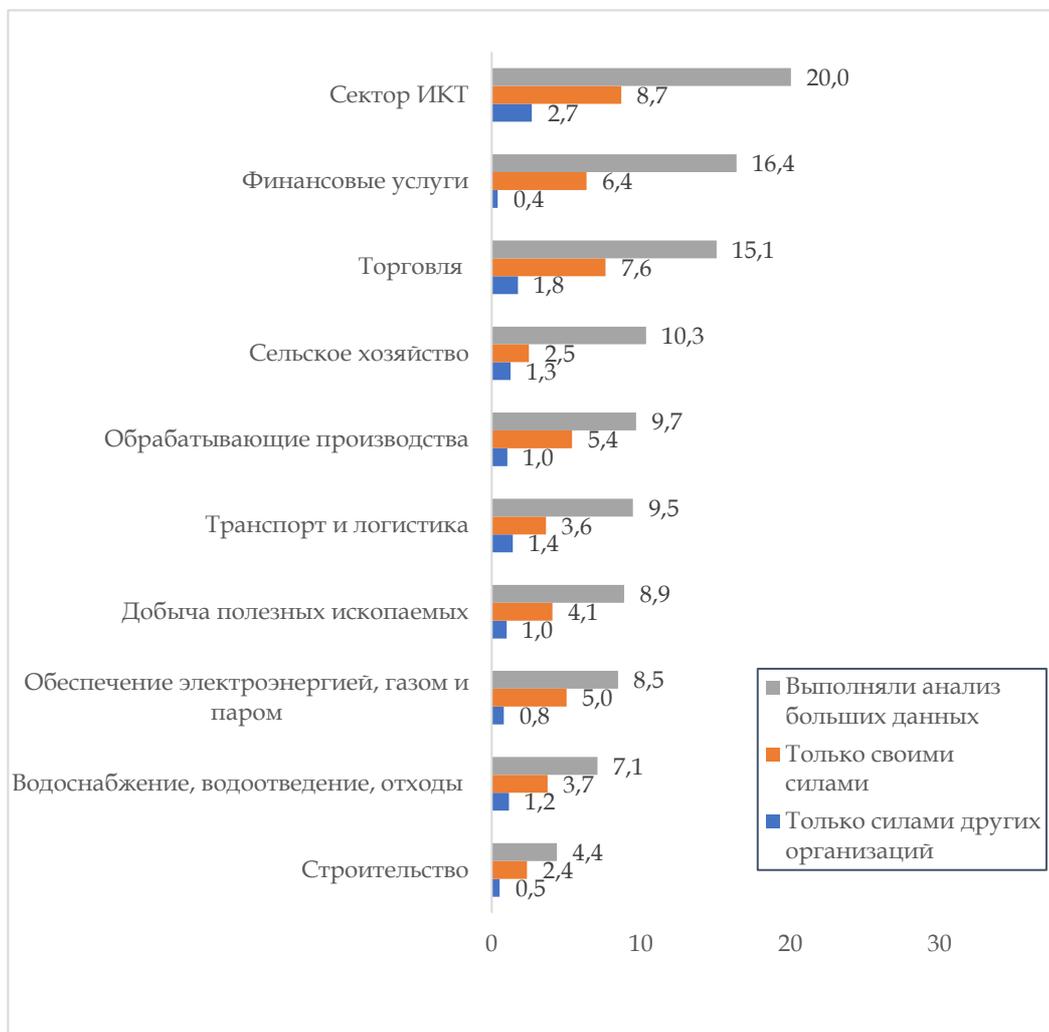


Рис. 3. Доля предприятий различных отраслей, выполнявших анализ больших данных, 2020, %

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.

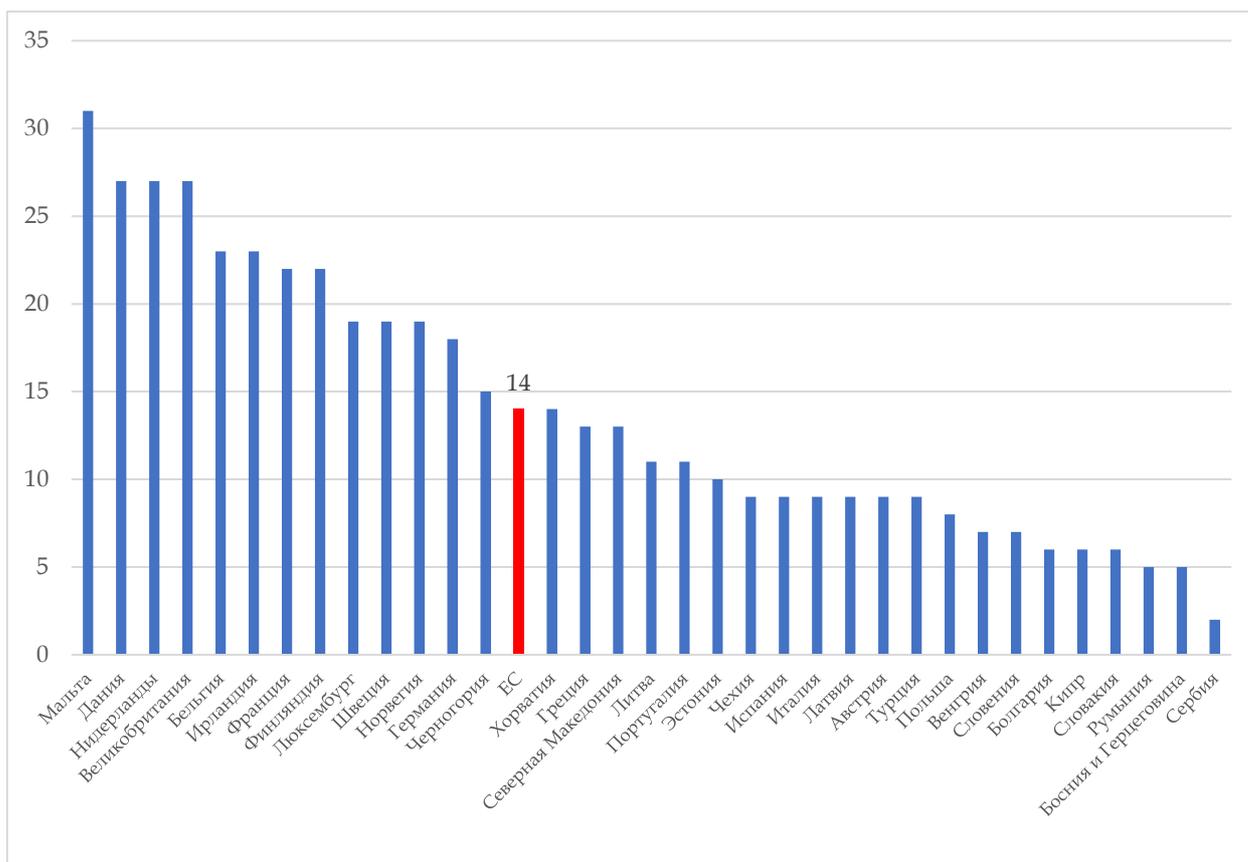


Рис. 4. Доля предприятий, проводящих анализ больших данных в странах ЕС, 2020, %
Источник: [27]

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)