

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](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.

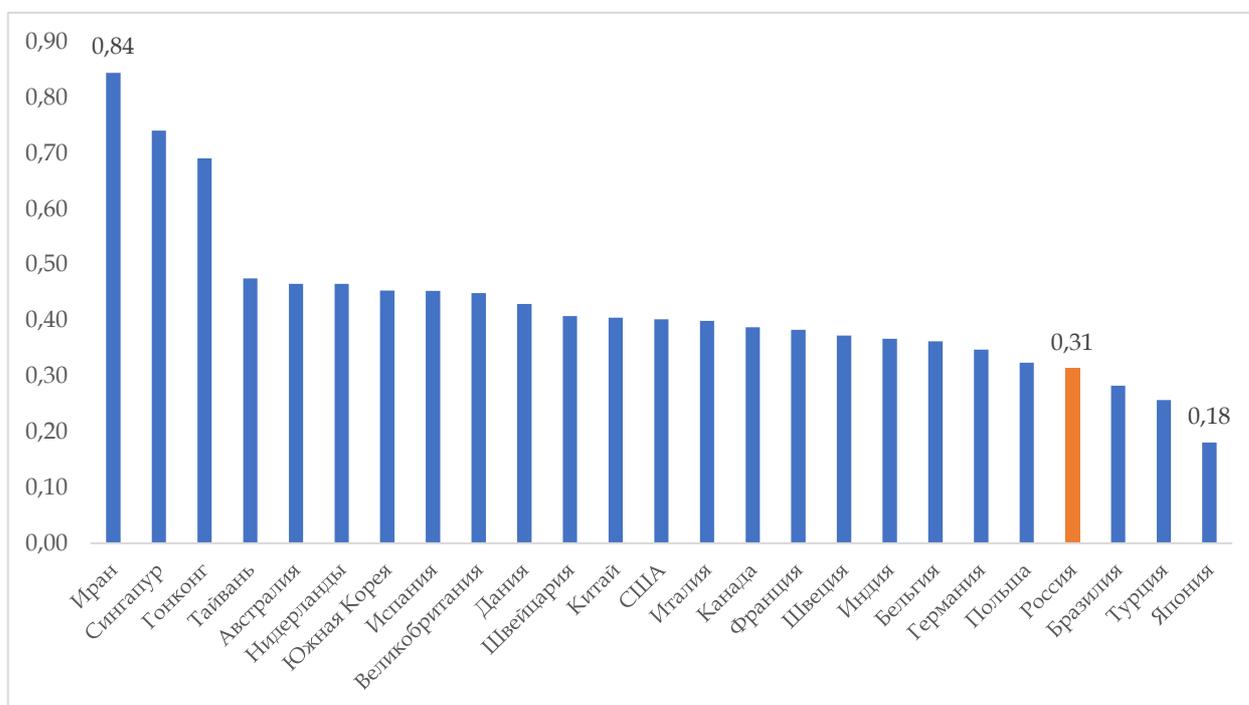


Рисунок 2. Доля публикаций с использованием технологий работы с большими данными (25 стран-лидеров по числу публикаций с большими данными, все области науки кроме Computer and information sciences и Electrical and electronic engineering, WoS, 2019 г.), %

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).

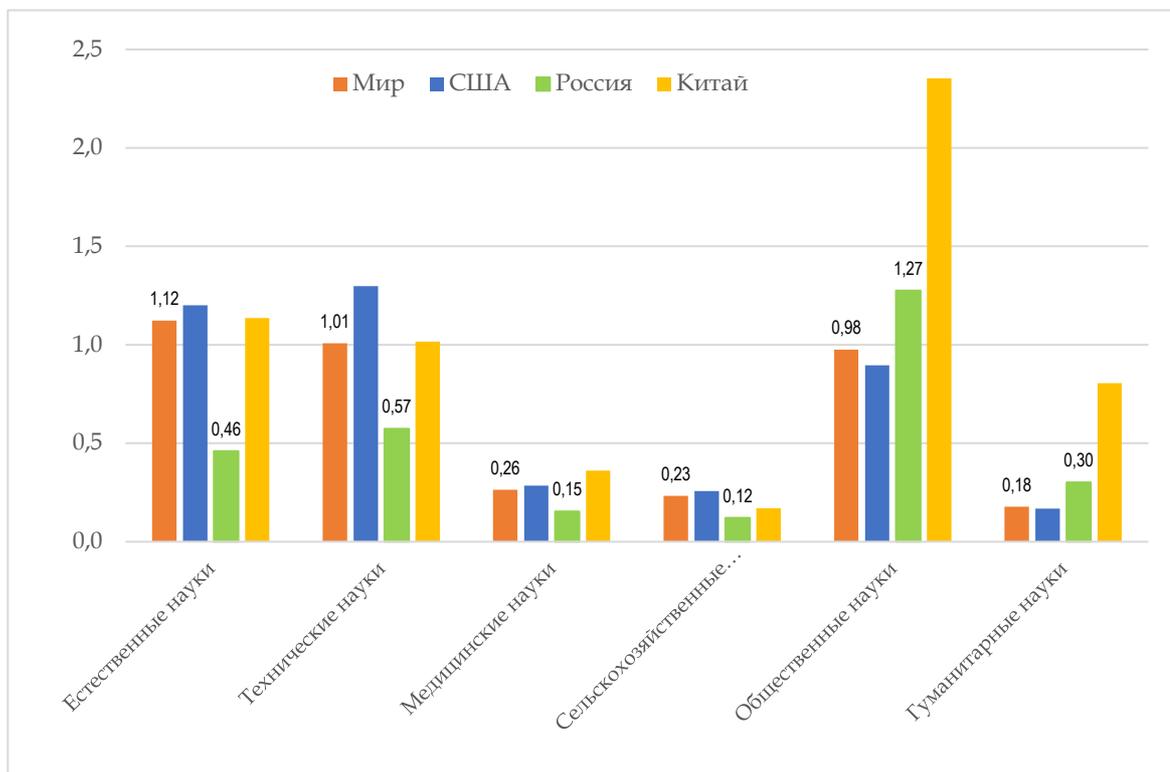


Рисунок 3. Доля публикаций с использованием технологий работы с большими данными, в различных областях науки (WoS, 2019 г.), %

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).

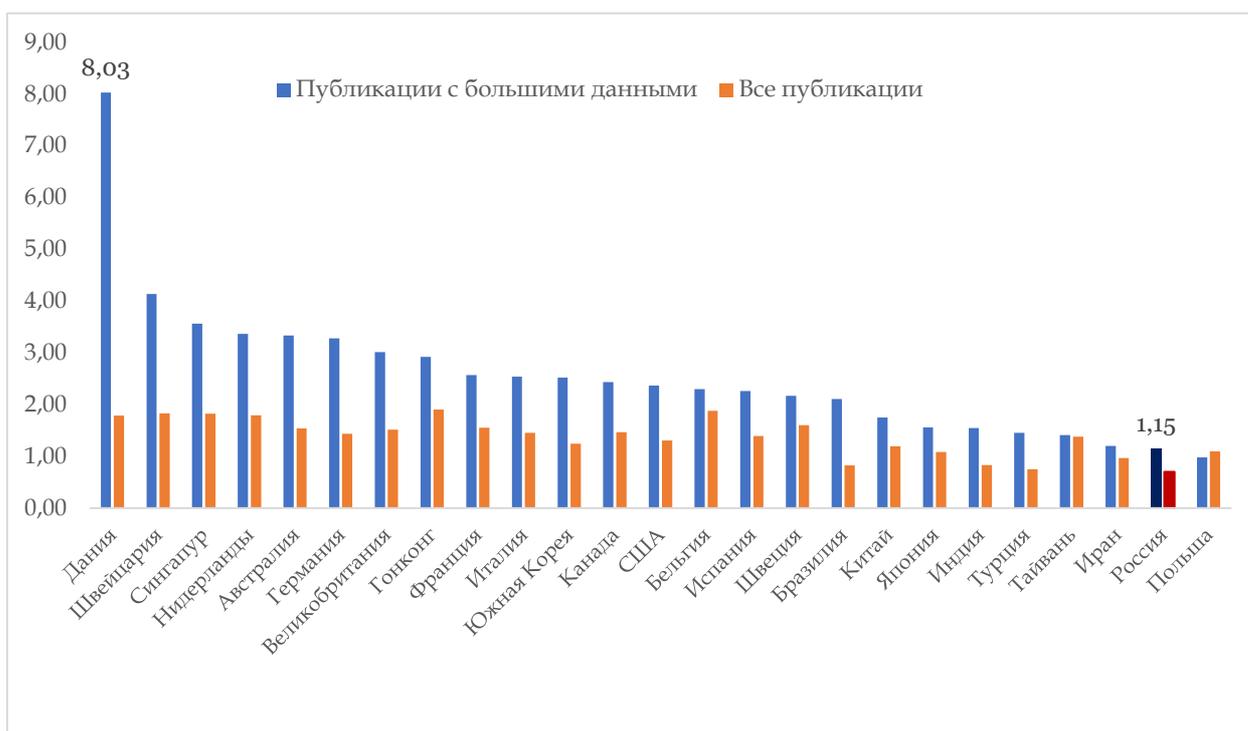


Рисунок 5. Средняя нормализованная цитируемость публикаций с использованием технологий работы с большими данными и всех публикаций (25 стран-лидеров по числу публикаций с большими данными, все области науки, кроме Computer and information sciences и Electrical and electronic engineering, WoS, 2019 г.), %

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.

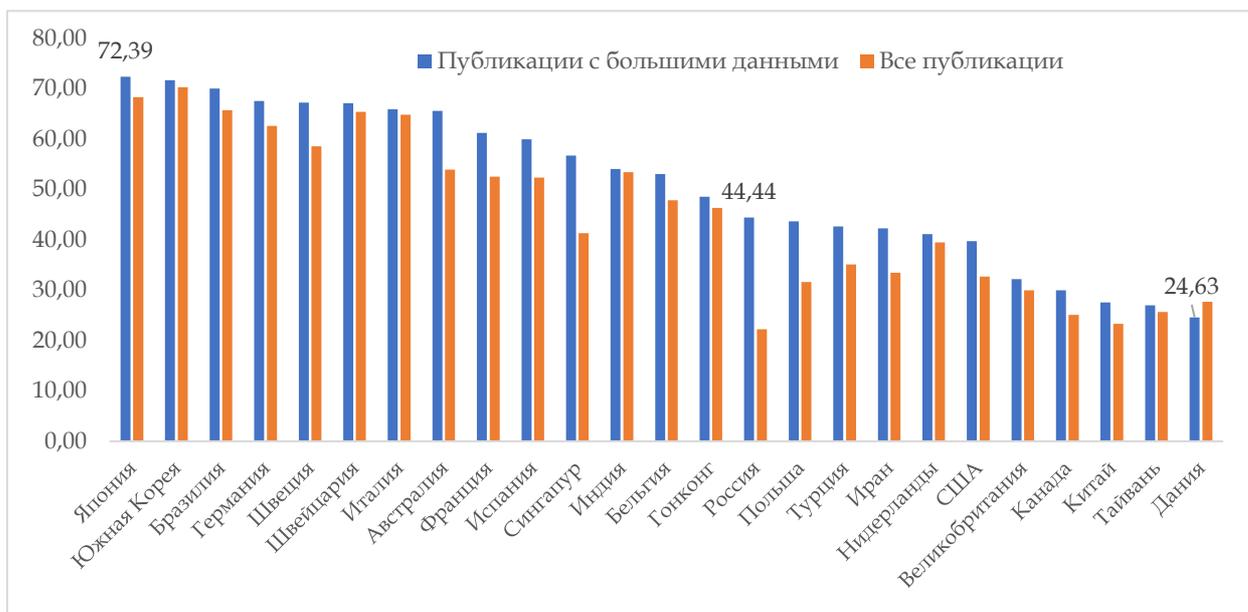


Рисунок 6. Международное соавторство публикаций с использованием технологий работы с большими данными и всех публикаций (25 стран-лидеров по числу публикаций с большими данными, все области науки, кроме Computer and information sciences и Electrical and electronic engineering, WoS, 2019 г.), %

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](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](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](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](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](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](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](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](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](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](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.