

# РАБОТА С ДАННЫМИ — ОСНОВА ЦИФРОВОЙ ЭКОНОМИКИ

## WORKING WITH DATA IS THE BASIS OF DIGITAL ECONOMY



### **ЕРШОВА ТАТЬЯНА ВИКТОРОВНА**

Директор Национального центра цифровой экономики и Центра компетенций НТИ по технологиям хранения и анализа больших данных, МГУ имени М.В. Ломоносова

### **TATIANA V. ERSHOVA**

Director of the National Center for Digital Economy and NTI Competence Center for Big Data Storage and Analytics, Lomonosov Moscow State University

### **АННОТАЦИЯ**

В статье подчеркивается роль данных для развития цифровой экономики в контексте положений доклада ЮНКТАД о цифровой экономике 2019 года и кратко описывается система управления данными на государственном уровне в Российской Федерации. Особое место отводится представлению созданного в МГУ в 2018 году Центра компетенций Национальной технологической инициативы по технологиям хранения и анализа больших данных, обеспечивающего разработку

инструментов эффективной работы с данными в целях максимального использования возможностей цифровой экономики и координацию деятельности в этой области на национальном уровне.

### **ABSTRACT**

The article emphasizes the role of data for the development of the digital economy in the context of the provisions of the UNCTAD Digital Economy Report 2019 and briefly describes the data management system at the state level in the Russian Federation. A special attention is paid to the presentation of the Competence Center of the National Technological Initiative for Big Data Storage and Analytics created in MSU in 2018. The center is providing for and coordinating activities related to the development of tools for efficient work with data in order to maximize the opportunities of the digital economy.

### **КЛЮЧЕВЫЕ СЛОВА**

Цифровая экономика, данные, большие данные, Национальная технологическая инициатива, управление данными, технологии хранения и анализа больших данных, «сквозные» технологии, мониторинг.

### **KEY WORDS**

Digital economy, data, big data, National Technological Initiative, data management, data storage and analytics, «cross-cutting» technologies, monitoring.

### **ВВЕДЕНИЕ**

Поводом к написанию этой статьи явилось событие, произошедшее 4 сентября 2019 года в штаб-квартире Международного Союза экономистов и Вольного экономического общества России, а именно официальное представление «Доклада о цифровой экономике 2019» Конференции ООН по торговле и развитию (ЮНКТАД), ранее известного как Доклад об информационной экономике. Важным тези-

сом обсуждаемого во всем мире документа является то, что «расширение цифровой экономики обусловлено цифровыми данными. Ведь цифровая экономика продолжает развиваться с головокружительной скоростью благодаря возможности собирать, использовать и анализировать огромные объемы машиночитаемой информации (цифровых данных) практически обо всем. Эти цифровые данные возникают из цифровых следов личной, социальной и деловой активности, осуществляемой на различных цифровых платформах» [1, с. XV]. Поэтому важное значение приобретают управление данными и создание эффективных механизмов сбора, хранения и анализа данных.

### **УПРАВЛЕНИЕ ДАННЫМИ В РОССИЙСКОЙ ФЕДЕРАЦИИ**

В России работе с данными уделяется существенное внимание на государственном уровне. Политика управления данными реализуется в соответствии с основными стратегическими документами [2, 3] и законами [4–6]. Управление конкретными видами данных, например, записями актов гражданского состояния или пространственными данными, как правило, регулируется отдельными стратегическими документами и законодательными актами [7–9].

Решения федеральных органов власти рекомендуются для использования региональными органами в рамках их полномочий.

Некоторые виды данных собираются на национальном уровне неправительственными организациями и используются органами власти, негосударственными организациями и гражданами. В качестве примеров можно привести Реестр сведений о полисах обязательного страхования, который ведет Российский союз автостраховщиков, или Единая инфор-

мационная система нотариата, которую ведет Федеральная нотариальная палата на основании соответствующих законодательных актов [10, 11].

Ключевым элементом российской государственной инфраструктуры работы с данными является Единая система межведомственного электронного взаимодействия (СМЭВ), созданная в рамках инфраструктуры электронного правительства Российской Федерации и активно используемая для обмена данными между органами государственной власти и местного самоуправления. Часть сервисов СМЭВ также предоставляется и негосударственным организациям, в частности, банкам. Достаточно подробный обзор состояния инфраструктуры работы с данными в Российской Федерации приведен в докладе «Цифровое правительство 2020: Перспективы для России» [12].

Дальнейшее развитие государственная политика управления данными в Российской Федерации получила в проекте «Национальная система управления данными» в рамках национальной программы «Цифровая экономика Российской Федерации». Разработана и утверждена Концепция создания и функционирования Национальной системы управления данными [13].

Что касается обеспечения безопасности данных, то соответствующий анализ российской политики в этой области приведен в докладе Института развития информационного общества и других российских организаций, подготовленном при поддержке Всемирного банка [14]. Следует особо отметить, что к оператору, осуществляющему обработку персональных данных, предъявляются обязательные требования по необходимым организационным и техническим мерам защиты [15].

Государственная политика использования персональных данных граждан закреплена в соответствующем законе [16], который был принят после присоединения России в 2001 г. к Конвенции Совета Европы о защите физических лиц при автоматизированной обработке персональных данных [17] и ее ратификации в 2016 г. В 2018 г. Российская Федерация подписала Протокол о модернизации Конвенции с целью повышения уровня защиты персональных данных на международном уровне. Для российских граждан это расширит права на получение информации о несанкционированном доступе третьих лиц к их персональным данным, а к российским компаниям, работающим на рынках ЕС, европейские регуляторы в рамках Общего регламента по защите данных [18] не будут применять дополнительные требования по мерам защиты персональных данных своих граждан.

В то же время в Российской Федерации лишь частично урегулированы вопросы информирования пользователей о том, как собираются, хранятся, обрабатываются и/или передаются их персональные данные. Например, в нормативных актах Российской Федерации в настоящее время не определены и не закреплены процедуры обезличивания персональных данных. На решение этой проблемы направлен проект «Цифровой профиль» в рамках национальной программы «Цифровая экономика Российской Федерации», концепция которого была утверждена в марте 2019 г. [19]. Уже начат эксперимент по повышению качества и связанности данных, содержащихся в государственных информационных ресурсах.

В российском законодательстве явно не установлен порядок и ответственность оператора персональных данных

за инциденты, связанные с утечкой персональных данных. В подобных случаях на практике к операторам персональных данных применяются штрафные санкции за нарушения в соответствии с Кодексом об административных правонарушениях [20]. При этом оператор персональных данных не обязан уведомлять пользователя о нарушении, допущенном при обработке его персональных данных, но обязан сообщить ему об этом в случае получения соответствующего запроса.

Имеются специальные категории персональных данных, которые органы власти имеют право собирать и использовать при определенных условиях без согласия субъектов персональных данных [21]. К ним относятся данные, необходимые для защиты жизни, здоровья или иных жизненно важных интересов субъекта персональных данных; данные, необходимые для обеспечения обороны, безопасности, противодействия терроризму, оперативно-разыскной деятельности и другие категории.

### **Создание Центра компетенций НТИ по технологиям хранения и анализа больших данных**

Управление — лишь одна из сторон работы с данными в условиях развития цифровой экономики. Колоссальный рост производимых человечеством данных порождает существенные проблемы, связанные с тем, что большие массивы данных не могут быть использованы сами по себе — они требуют специализированных средств анализа и использования. В различных сферах деятельности приходится иметь дело с данными в самых разных форматах, в частности, с неструктурированными данными в виде документов, видео,

изображений и аудиофайлов и т.д. Это приводит к высокой сложности систем обработки и анализа данных.

Несовершенство большинства современных подходов связано с тем, что они сосредоточены на вопросах поиска, сбора, передачи и хранения данных. В результате накапливаются огромные массивы данных, подавляющая часть которых не нужна для решения задач, ради которых их собирали, и никогда не используется. Собираемые данные в некотором смысле аналогичны добываемым ресурсам: подобно тому, как природные ресурсы добывают только с определенной целью, данные также нужно собирать, предварительно определив, зачем они нужны.

Понимание этого послужило импульсом к участию Московского университета в конкурсе на государственную поддержку центров Национальной технологической инициативы по так называемым «сквозным» технологиям на базе организаций высшего образования и научных организаций, объявленном Российской венчурной компанией в начале 2018 года. Благодаря победе в этом конкурсе по направлению «Технологии хранения и анализа больших данных» в МГУ в структуре Национального центра цифровой экономики был создан Центр хранения и анализа больших данных.

Сущность предлагаемого центром подхода к анализу больших данных заключается в том, что он ставит во главу угла цели, а не средства. Уже после определения целей выбираются требования к данным, средства их сбора и анализа. В результате правильного выбора пропадает необходимость в хранении огромных объемов избыточных, никогда не используемых данных, снижаются требования к объему хранилища данных и значительно снижаются объемы потоков информации. Данный подход являет-

ся комплексным. Он предполагает участие специалистов в различных предметных областях, которые помогут определить истинные цели собираемых данных.

Программа развития центра на 2019 год включает пять блоков:

- I. Реализация ключевых комплексных научно-исследовательских и опытно-конструкторских проектов.
- II. Обеспечение правовой охраны, управления правами и защиты результатов интеллектуальной деятельности (РИД).
- III. Разработка и реализация основных образовательных программ высшего образования, программ дополнительного образования, дисциплин (модулей), направленных на формирование компетенций.
- IV. Развитие информационной инфраструктуры, а также инфраструктуры научной, научно-технической и инновационной деятельности.
- V. Развитие партнерских отношений.

Задачи научно-исследовательского направления группируются вокруг восьми ключевых проектов, из которых два первых являются приоритетными:

1. Предиктивная аналитика технических систем.
2. Система автоматического поиска уязвимостей в веб-приложениях.
3. Облачные технологии обработки и интерпретации медицинских диагностических изображений на основе применения средств анализа больших данных.
4. Средства интеллектуального анализа больших массивов текстов.
5. Математические основы интеллектуального анализа больших данных.



6. Новые подходы к проектированию систем считывания для технологии сверхстабильной трехмерной оптической памяти с многоуровневым кодированием на носителях из стекла.
7. Интеллектуальный анализ больших данных в задачах экологии и охраны окружающей среды.
8. Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации.

Последний (по порядку перечисления, но не по значимости) проект представляется весьма важным для функционирования не только описываемого центра, но и всех центров компетенций НТИ по «сквозным» технологиям. В результате его выполнения будет сформирована комплексная система измеримых показателей, инструментарий и методика проведения обследований, которая позволит объективно оценивать состояние развития технологий хранения и анализа больших данных. Будет разработан онлайн-инструмент для проведения самооценки уровня готовности к использованию этих технологий предприятиями и организациями. Наряду с этим в ближайшие годы будет проведена работа по модернизации серии международных стандартов по технологиям хранения и анализа больших данных и утверждению их в качестве национальных стандартов. Подходы и результаты проекта могут быть использованы и другими центрами компетенций НТИ по «сквозным» цифровым технологиям.

Реализации этого проекта предшествовало активное участие в 2018 году сотрудников Центра компетенций НТИ по технологиям хранения и анализа больших данных в подготовке пилотной версии Национального индекса разви-

тия цифровой экономики [22] под эгидой госкорпорации «Росатом» как центра компетенций федерального проекта «Цифровые технологии» национальной программы «Цифровая экономика Российской Федерации». В этой работе проанализированы факторы, влияющие на развитие цифровой экономики, как «цифровые», так и «нецифровые». «Цифровыми» факторами являются соответствующая инфраструктура сетей передачи данных, центров обработки данных и цифровых платформ. «Нецифровые» факторы — это человеческий капитал, НИОКР и инновации, деловая среда, государственная политика и регулирование, информационная безопасность. Особым фактором, влияющим на развитие цифровой экономики, является наличие в стране зрелого цифрового сектора экономики, включающего сектор информационно-коммуникационных технологий наряду с сектором (цифрового) контента и СМИ.

Кроме того, при подготовке индекса был проведен анализ использования цифровых технологий в сферах государственного управления, здравоохранения, бизнеса, а также их применение обычными гражданами (потребителями). Также был сделан краткий сравнительный анализ уровня использования цифровых технологий в 32 европейских странах и представлена «тепловая карта» цифровизации отдельных отраслей экономики в этих странах, обозначена необходимость анализа воздействия цифровой трансформации на развитие экономики.

Что касается работы Центра компетенций НТИ по технологиям хранения и анализа больших данных в области обеспечения правовой охраны, управления правами и защиты РИД, то она включает выявление и охрану технических решений, подготовку и подачу заявок на патенты, сопровожде-

ние лицензионных сделок и проектов, а также мероприятия, направленные на коммерциализацию.

Образовательная деятельность центра направлена на подготовку высококлассных специалистов в области хранения и анализа больших и сверхбольших данных. Планируется создание обновленной образовательной платформы на базе существующих магистерских программ, специализированных курсов и лабораторных практикумов, а также разработка новых специализированных, углубленных курсов по отдельным направлениям, электронных учебных модулей.

В рамках разрабатываемой образовательной платформы ведется разработка курсов лекций по следующим направлениям: аналитика больших данных, хранение больших данных, программная инженерия и информационные технологии, предметно-ориентированные знания. Кроме того, предполагается организовать курсы повышения квалификации для студентов, аспирантов, молодых ученых, реализовать образовательные программы по нескольким направлениям, соответствующим смежным с технологиями хранения и анализа больших данных «сквозным» цифровым технологиям Национальной технологической инициативы. Предусмотрена разработка и реализация программы повышения квалификации для государственных служащих и работников госкомпаний по формированию навыков профессиональной работы с информационно-аналитическими и экспертно-аналитическими средствами поддержки принятия решений, а также навыков подготовки проектов технических заданий и финансово-экономических обоснований работ по модернизации действующих и созданию новых ситуационных центров.

В качестве целевых потребителей результатов НИР и НИОКР, выполненных в рамках реализации программы

центра, а также разработанных учебных программ являются средние и крупные корпоративные заказчики, органы государственной власти, научно-исследовательские, образовательные и медицинские учреждения, организации сферы культуры.

Основой инфраструктуры центра является суперкомпьютерный комплекс МГУ имени М.В. Ломоносова — крупнейший российский комплекс с высшей для России позицией в рейтинге самых быстрых компьютеров мира. Возможностями этого комплекса пользуются свыше 2200 сотрудников МГУ и более чем 200 научных и учебных организаций России, решающих ежедневно около 1000 задач. Будет осуществлено расширение ресурсных возможностей комплекса с целью выделения специальных ресурсов и квот для проектов, выполняемых в рамках программы центра.

В рамках развития партнерских отношений центра сформирован консорциум, в который на момент подготовки данной статьи входят 32 организации различных типов: 11 образовательных организаций высшего образования, 5 научных организаций, 1 государственная корпорация, 14 коммерческих компаний и 1 некоммерческая организация. Участники консорциума имеют большие заделы с точки зрения проведения фундаментальных и/или прикладных исследований в области хранения и анализа больших данных, разработки и внедрения крупных информационных систем, обладают значительной материально-технической базой и большим суммарным человеческим капиталом мирового уровня, занимают лидирующие либо весьма твердые позиции в своих сферах деятельности, проявляют большой интерес к долгосрочному сотрудничеству, нацелены на убедительные результаты реализации программы центра. На-

личие среди участников консорциума организаций разного масштаба (от крупных учреждений и компаний до небольших игроков рынка) способствует лучшему пониманию и более эффективной реализации целей и задач программы центра.

Более подробную информацию о деятельности центра можно получить на его сайте [23].

### **Библиографический список**

1. Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries / United Nations Conference on Trade and Development. N.Y.: UN, 2019. 172 p.
2. Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы. Утверждена Указом Президента Российской Федерации от 09.05.2017 № 203.
3. Доктрина информационной безопасности Российской Федерации. Утверждена Указом Президента Российской Федерации от 05.12.2016 № 646.
4. Концепция создания и функционирования национальной системы управления данными. Утверждена распоряжением Правительства Российской Федерации от 03.06.2019 № 1189-р.
5. Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ (ред. от 01.05.2019).
6. Федеральный закон «О персональных данных» от 27.07.2006 № 152-ФЗ (ред. от 31.12.2017).
7. Федеральный закон «Об актах гражданского состояния» от 15.11.1997 № 143-ФЗ (ред. от 29.05.2019).
8. Концепция создания и развития инфраструктуры пространственных данных Российской Федерации. Утверждена распоряжением Правительства РФ от 21.08.2006 № 1157-р.

9. Федеральный закон «О геодезии, картографии и пространственных данных и о внесении изменений в отдельные законодательные акты Российской Федерации» от 30.12.2015 № 431 (ред. от 03.08.2018).
10. Федеральный закон «Об обязательном страховании гражданской ответственности владельцев транспортных средств» от 25.04.2002 № 40-ФЗ (ред. от 01.05.2019).
11. Основы законодательства Российской Федерации о нотариате» (утв. ВС РФ 11.02.1993 № 4462-1) (ред. от 27.12.2018).
12. Petrov Oleg V.; Bunchuk Mikhail; Stott Andrew Charles; Hohlov Yuri. 2016. Digital Government 2020: Prospects for Russia. Washington, D.C.: World Bank Group. URL: <http://documents.worldbank.org/curated/en/2016/05/26367056/digital-government-2020-prospects-russia> (последнее обращение 29.09.2019).
13. Концепция создания и функционирования национальной системы управления данными. Утверждена распоряжением Правительства Российской Федерации от 03.06.2019 № 1189-р.
14. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. 166 с.
15. Федеральный закон «О персональных данных» от 27.07.2006 № 152-ФЗ (ред. от 31.12.2017), ст. 19, п. 1.
16. Федеральный закон «О персональных данных» от 27.07. 2006 № 152-ФЗ (ред. от 31.12.2017).
17. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Council of Europe. European Treaty Series — No. 108. Strasbourg, 28.01.1981. URL: <https://rm.coe.int/1680078b37> (последнее обращение 30.09.2019).
18. General Data Protection Regulation — Регламент ЕС 2016/679 от 27.04.2016. URL: [https://ec.europa.eu/info/law/law-topic/data-protection\\_en](https://ec.europa.eu/info/law/law-topic/data-protection_en) (последнее обращение 30.09.2019).

19. Концепция и архитектура цифрового профиля — ЕСИА 2.0 / Банк России и Минкомсвязь России. — 87 с. (Утв. решением президиума Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности 29.03.2019).
20. Кодекс Российской Федерации об административных правонарушениях» от 30.12.2001 № 195-ФЗ (ред. от 17.06.2019), ст. 13.11.
21. Федеральный закон «О персональных данных» от 27.07. 2006 № 152-ФЗ (ред. от 31.12.2017), ст. 10.
22. Национальный индекс развития цифровой экономики: Пилотная реализация. М: Госкорпорация «Росатом», 2018. 92 с.
23. Центр хранения и анализа больших данных МГУ имени М.В. Ломоносова. URL: <https://bigdata-msu.ru> (последнее обращение 26.09.2019).

## References

1. Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries / United Nations Conference on Trade and Development. N.Y.: UN. 2019. 172 p.
2. Strategiya razvitiya informatsionnogo obshchestva v Rossiyskoy Federatsii na 2017 — 2030 gody. Utverzhdena Ukazom Prezidenta Rossiyskoy Federatsii ot 09.05.2017 № 203.
3. Doktrina informatsionnoy bezopasnosti Rossiyskoy Federatsii. Utverzhdena Ukazom Prezidenta Rossiyskoy Federatsii ot 05.12.2016 № 646.
4. Kontseptsiya sozdaniya i funktsionirovaniya natsionalnoy sistemy upravleniya dannymi. Utverzhdena rasporyazheniyem Pravitelstva Rossiyskoy Federatsii ot 03.06.2019 № 1189-r.
5. Federalnyy zakon «Ob informatsii, informatsionnykh tekhnologiyakh i o zashchite informatsii» ot 27.07.2006 № 149-FZ (red. ot 01.05.2019).

6. Federalnyy zakon «O personalnykh dannykh» ot 27.07.2006 № 152-FZ (red. ot 31.12.2017).
7. Federalnyy zakon «Ob aktakh grazhdanskogo sostoyaniya» ot 15.11.1997 № 143-FZ (red. ot 29.05.2019).
8. Kontseptsiya sozdaniya i razvitiya infrastruktury prostranstvennykh dannykh Rossiyskoy Federatsii. Utverzhdena rasporyazheniyem Pravitelstva RF ot 21.08.2006 № 1157-r.
9. Federalnyy zakon «O geodezii, kartografii i prostranstvennykh dannykh i o vnesenii izmeneniy v otdelnyye zakonodatelnyye akty Rossiyskoy Federatsii» ot 30.12.2015 № 431 (red. ot 03.08.2018).
10. Federalnyy zakon «Ob obyazatelnom strakhovanii grazhdanskoy otvetstvennosti vladeltsev transportnykh sredstv» ot 25.04.2002 № 40-FZ (red. ot 01.05.2019).
11. Osnovy zakonodatelstva Rossiyskoy Federatsii o notariate» (utv. VS RF 11.02.1993 № 4462-1) (red. ot 27.12.2018).
12. Petrov. Oleg V.; Bunchuk. Mikhail; Stott. Andrew Charles; Hohlov. Yuri. 2016. Digital Government 2020: Prospects for Russia. Washington, D.C.: World Bank Group.
13. URL: <http://documents.worldbank.org/curated/en/2016/05/26367056/digital-government-2020-prospects-russia> (posledneye obrashcheniye 29.09.2019).
14. Kontseptsiya sozdaniya i funktsionirovaniya natsionalnoy sistemy upravleniya dannymi. Utverzhdena rasporyazheniyem Pravitelstva Rossiyskoy Federatsii ot 03.06.2019 № 1189-r.
15. Analiz tekushchego sostoyaniya razvitiya tsifrovoy ekonomiki v Rossii. M.: Institut razvitiya informatsionnogo obshchestva. 2018. 166 s.
16. Federalnyy zakon «O personalnykh dannykh» ot 27.07.2006 № 152-FZ (red. ot 31.12.2017), st. 19, p. 1.
17. Federalnyy zakon «O personalnykh dannykh» ot 27.07. 2006 № 152-FZ (red. ot 31.12.2017).



18. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Council of Europe. European Treaty Series — No. 108. Strasbourg. 28.01.1981. URL: <https://rm.coe.int/1680078b37> (posledneye obrashcheniye 30.09.2019).
19. General Data Protection Regulation — Reglament ES 2016/679 ot 27.04.2016. URL: [https://ec.europa.eu/info/law/law-topic/data-protection\\_en](https://ec.europa.eu/info/law/law-topic/data-protection_en) (posledneye obrashcheniye 30.09.2019).
20. Kontsepsiya i arkhitektura tsifrovogo profilya — ESIA 2.0 / Bank Rossii i Minkomsvyaz Rossii. — 87 s. (Utv. resheniyem prezidiuma Pravitelstvennoy komissii po tsifrovomu razvitiyu. ispolzovaniyu informatsionnykh tekhnologiy dlya uluchsheniya kachestva zhizni i usloviy vedeniya predprinimatelskoy deyatelnosti 29.03.2019).
21. Kodeks Rossiyskoy Federatsii ob administrativnykh pravonarusheni-yakh» ot 30.12.2001 № 195-FZ (red. ot 17.06.2019). st. 13.11.
22. Federalnyy zakon «O personalnykh dannykh» ot 27.07. 2006 № 152-FZ (red. ot 31.12.2017). st. 10.
23. Natsionalnyy indeks razvitiya tsifrovoy ekonomiki: Pilotnaya realizatsiya. M: Goskorporatsiya «Rosatom». 2018. 92 s.
24. Tsentr khraneniya i analiza bolshikh dannykh MGU imeni M.V. Lomonosova. URL: <https://bigdata-msu.ru/> (posledneye obrashcheniye 26.09.2019).