

Инвестиционная и инновационная деятельность

д-р экон. наук, профессор
Ю.И. Селиверстов
Белгородский государственный
технологический университет
им. В.Г. Шухова

ВКЛАД ТЕХНОЛОГИИ BIG DATA В ЭКОНОМИЧЕСКОЕ И ИННОВАЦИОННОЕ РАЗВИТИЕ РОССИИ

Развитию информационных технологий в значительной мере препятствовало отсутствие возможности хранения и обработки больших объемов информации, структурированных и неструктурированных данных. Перелом произошел в конце первого десятилетия XXI века, когда была сформулирована проблема извлечения знаний, доступных для восприятия человеком из больших массивов цифровой информации.

Впервые термин «большие данные» появился в прессе в 2008 году, когда редактор журнала Nature Клиффорд Линч выпустил статью на тему развития будущего науки с помощью технологий работы с большим количеством данных. До 2009 года данный термин рассматривался только с точки зрения научного анализа, но после выхода еще нескольких статей пресса стала широко использовать понятие Big Data – и продолжает использовать его в настоящее время. В 2010 году стали появляться первые попытки решить нарастающую проблему больших данных. Были выпущены программные продукты, действие которых было направлено на то, чтобы минимизировать риски при использовании огромных информационных массивов. К 2011 году большими данными заинтересовались такие крупные компании, как Microsoft, Oracle, EMC и IBM – они стали первыми использовать наработки Big Data в своих стратегиях развития, причем довольно успешно. [7]

В настоящее время термин «большие данные» не имеет общепринятого определения. В зависимости от источника происхождения большие данные можно условно разделить на две большие группы:

а) *промышленные «большие данные»*, которые собираются с помощью различного рода сенсоров, датчиков устройств, используемых в промышленности, сельском хозяйстве и отображают показатели таких устройств (эти данные обычно выводятся в категорию интернета вещей);

б) *большие пользовательские данные* создаются в процессе использования различных приложений, сервисов в сети Интернет (данные профай-

лов в социальных сетях, геолокационные данные, информация о пользовательском поведении и т.д.).

В литературе можно встретить различные определения термина «большие данные». Так, например, эксперты Сретенского клуба им. С.П. Курдюмова определяют *большие данные (Big Data)* как совокупность подходов, инструментов и методов, предназначенных для обработки структурированных и неструктурированных данных (в т. ч. из разных независимых источников) с целью получения воспринимаемых человеком результатов. Большие данные характеризуются значительным объемом, разнообразием и скоростью обновления, что делает стандартные методы и инструменты работы с информацией недостаточно эффективным. Таким образом, технология Больших данных – это инструмент принятия решений на основе больших объемов информации [1].

Исследователи Высшей школы экономики дают следующее определение. *Большие данные* – технологии сбора, обработки и хранения структурированных и неструктурированных массивов информации, характеризующихся значительным объемом и быстрой скоростью изменений (в т. ч. в режиме реального времени), что требует специальных инструментов и методов работы с ними [3].

В.Д. Маркова определяет Big Data как многоплановое и многоуровневое понятие, основными характеристиками которого являются объемы данных, скорость их генерации и доступности, разнообразие источников и типов данных, форм их хранения. При этом чаще всего речь идет о неструктурированных или плохо структурированных генерируемых данных, которые надо интерпретировать в понятных для пользователя терминах [2].

Таким образом, Big Data – это не только сами данные, но и технологии их обработки и использования, методы поиска необходимой информации в больших массивах. Проблема больших данных по-прежнему остается открытой и жизненно важной для любых систем, десятилетиями накапливающих самую разнообразную информацию [7].

В современном мире Big Data – социально-экономический феномен, который связан с тем, что появились новые технологические возможности для анализа огромного количества данных. Огромные объемы данных обрабатываются для того, чтобы человек мог получить конкретные и нужные ему результаты для их дальнейшего эффективного применения.

По данным компании IBS, к 2003 году мир накопил 5 эксабайтов данных (1 ЭБ = 1 млрд гигабайтов). К 2008 году этот объем вырос до 0,18 зеттабайта (1 ЗБ = 1024 эксабайта), к 2011 году – до 1,76 зеттабайта, к 2013 году – до 4,4 зеттабайта. В мае 2015 года глобальное количество данных превысило 6,5 зеттабайта. К 2020 году, по прогнозам, человечество сформи-

мирует 40-44 зеттабайтов информации. А к 2025 году вырастет в 10 раз, говорится в докладе The Data Age 2025, который был подготовлен аналитиками компании IDC. В докладе отмечается, что большую часть данных генерировать будут сами предприятия, а не обычные потребители. Аналитики исследования считают, что данные станут жизненно-важным активом, а безопасность – критически важным фундаментом в жизни. Также авторы работы уверены, что технология изменит экономический ландшафт, а обычный пользователь будет коммуницировать с подключёнными устройствами около 4800 раз в день [5]. Big Data предоставляет пользователям следующие преимущества [6]:

1. Сбор данных из разных источников.
2. Улучшение бизнес-процессов через аналитику в реальном времени.
3. Хранение огромного объема данных.
4. Инсайты. Big Data более проникательна к скрытой информации при помощи структурированных и полуструктурированных данных.
5. Большие данные помогают уменьшать риск и принимать умные решения благодаря подходящей риск-аналитике.

Безусловно Big Data должна являться одним из основных приоритетов при разработке программ развития цифровой экономики. Обладателями данных для Big Data объективно являются социальные сети, крупные банки, операторы связи, федеральные торговые сети. Они достаточно успешно используют большие данные для коммерческих целей. Государство также имеет огромный ресурс больших данных самого разного плана (государственные архивы, статистика, транзакции в государственных реестрах, данные о товарах и услугах на уровне налоговой инспекции и т.п.).

Согласно прогнозу Bain & Company's Insights Analysis, к 2020 году наибольшую часть мирового рынка Big Data, по-прежнему, будет занимать финансовая отрасль, чуть меньше на эту сферу будут тратить интернет-компании, госсектор и телекоммуникации, а вот самым быстрорастущим сектором станет энергетика. При этом в Heavy Reading 2020 год называют рубежом, за которым операционная эффективность должна стать второй по значимости сферой приложения Big Data после улучшения клиентского сервиса, опередив точечный маркетинг [4].

По оценкам экспертов, в настоящее время российский рынок «больших данных» оценивается в сумму от 10 млрд рублей до 30 млрд рублей. Предполагается, что эффективное использование «больших данных» и внедрение технологии искусственного интеллекта во все сферы экономики обеспечит дополнительный рост ВВП на 1,5% ежегодно к 2024 году.

«Большие данные» входят в структуру сквозных цифровых технологий федерального проекта «Цифровые технологии».

ООО «Национальный центр информатизации» разработал проект «дорожной карты» развития технологии «больших данных». Документ разделяет данную технологию на шесть субтехнологий:

- 1) сбора данных;
- 2) хранения данных;
- 3) обработки и управления данными;
- 4) обогащения данных;
- 5) вывода данных;
- 6) предиктивная аналитика.

План мероприятий по развитию технологии «большие данные» состоит из 48 мероприятий и рассчитан до 2024 года. Документ формирует план совместных действий бизнес-сообщества и федеральных и региональных органов исполнительной власти по разработке и применению анализа больших массивов данных. По прогнозным оценкам, инвестиционная потребность «больших данных» к 2024 году составит 195 млрд рублей, в том числе по 60 млрд рублей потребуется для субтехнологий сбора данных интернета вещей и конвергентных программно-определяемых хранилищ (табл. 1).

Таблица 1

**Оценка инвестиционной потребности российских субтехнологий
«больших данных»**

	2020	2021	2022	2023	2024
Технологии сбора данных интернета вещей, млрд рублей	25	30	40	50	60
Конвергентные программно-определяемые хранилища данных, млрд рублей	45	50	55	60	60
Бизнес-аналитика (Business Intelligence, BI), млрд рублей	10	12	14	15	15
Обработка и утилизация данных с использованием AI и машинного обучения, млрд рублей	20	25	30	30	30
Обогащение данных, млрд рублей	10	12	14	14	15
Предиктивная аналитика, млрд рублей	10	12	14	15	15

Ожидается, что к 2024 году за счет «больших данных» отрасли добычи полезных ископаемых получат дополнительно 430 млрд рублей, сферы торговли и ремонта – 878 млрд рублей, недвижимости – 252 млрд рублей, строительства – 480 млрд рублей, транспортировки и хранения – 235 млрд рублей, финансов и страхования – 141 млрд рублей, здравоохранения и социальных услуг – 314 млрд рублей, обрабатывающие производства – 408 млрд рублей, связь и ИТ-отрасли – 81 млрд рублей.

В сфере *административной деятельности* дополнительно будет получено 352 млрд рублей, государственного управления, обеспечения военной безопасности и социального обеспечения – 292 млрд рублей, профессиональной, научной и технической деятельности – 271 млрд рублей, электроэнергии, газа и пара – 54 млрд рублей.

В сфере *добычи полезных ископаемых* применение «больших данных» поможет в 2,5 раза повысить эффективность разведки полезных ископаемых, в 2 раза улучшить показатели промышленной безопасности. На 10% будет увеличена скорость бурения скважин, а «цифровые двойники» позволят спрогнозировать более 75% производственных аварий.

В сфере *торговли и ремонта* применение «больших данных» на 10% сократит потери при производстве, транспортировке и хранении товаров. 75% предложений будет формироваться на основе анализа потребительской активности, 25% предложений будет формироваться с учетом кросс-отраслевых рынков.

В сфере *недвижимости и строительства* половина жилищного строительства, 75% торгово-развлекательных объектов и все объекты транспортной инфраструктуры будут строиться при прогнозировании заселения, транспортных потоков, загруженности жилого района и торговой проходимости. На 10% будут снижены риски невозврата при использовании кредитных средств при строительстве, на 5% снижена стоимость строительства и до 50% снижены затраты на эксплуатацию построенных объектов.

В сфере *транспортировки и хранения* на 10% будет сокращено время пребывания в пути на городских маршрутах и на столько же увеличена утилизация грузового автомобильного транспорта. На 15% будет сокращено время доставки грузов железнодорожным транспортом.

В *финансовой и страховой сферах* 80% предложений будет формироваться на основе анализа потребительской активности, 75% кредитных предложений будут учитывать кросс-отраслевые скоринговые модели. Также на 25% будет сокращена выдача невозвратных кредитных продуктов.

В области *здравоохранения и социальных услуг* половина диагнозов в онкологии, сердечно-сосудистых заболеваниях и инфекционных заболеваниях будет ставиться с использованием больших данных, 15% лекарств будет назначаться с учетом персонифицированного подхода.

В области *государственного управления* половина принимаемых органами государственной власти решений будет основана на анализе «больших данных», половина статистических данных будет собираться с использованием альтернативных методов получения первичных данных. При этом 90% ГИС будет доступно для использования бизнесом.

В области *связи и ИТ-отрасли* 75% предложений будет формироваться на основе анализа перспективных технологий и такая же доля потребителей будет удовлетворена качеством услуг связи. Также 30% предложений будет учитывать кросс-платформенные потребительские предпочтения.

Однако, внедрение технологии Big Data и развитие рынка «больших данных» сдерживается целым рядом факторов.

1. Основным сдерживающим фактором, особенно в компаниях среднего и малого бизнеса, остается стратегия выживания, а не развития, и, как следствие, экономия на ИТ-бюджете.

2. Инструменты обработки «больших данных» требуют больших вычислительных мощностей, которые дороги в закупке, установке и использовании.

3. Наблюдается дефицит специалистов, которые должны реализовывать проекты в сфере использования «больших данных». В России пока не образованы центры компетенций, которые занимались бы массовой подготовкой кадров с соответствующими компетенциями.

4. Актуальна проблема большого массива разрозненных или недостоверных данных.

5. Несмотря на то, что инструменты киберзащиты нового поколения, активно внедряются в бизнес-практику, пользователи систем Big Data продолжают опасаться утечки конфиденциальной информации о компаниях, а также персональных данных клиентов.

Еще одним из факторов, тормозящих темпы роста рынка «больших данных» в России, является несовершенство законодательства о персональных данных. Информация о физических лицах все активнее используется компаниями в коммерческих целях. Это происходит, прежде всего, из-за интенсификации использования информационных технологий, поскольку в цифровом пространстве можно легко получить данные о человеке.

Капитализация многих компаний напрямую зависит от объема обрабатываемых ими персональных данных, так называемых, больших пользовательских данных. В результате коммерческого использования таких данных появился новый рынок с соответствующими участниками - биржи данных, информационные брокеры, которые на профессиональной основе собирают информацию о физических лицах и предоставляют ее заинтересованным компаниям. На данный момент такая практика деятельности находится вне рамок правового регулирования. В контексте защиты неприкосновенности частной жизни основные риски использования «больших данных» можно свести к утечке данных (т.е. утрате контроля за их использованием, попаданию их к нежелательным лицам и возможности

совершения мошеннических действий с их помощью), а также к обработке неточных или неполных данных, наносящей вред правам граждан.

Следует отметить, что в сфере правового регулирования потоков данных в цифровом пространстве в нашей стране до сих пор не разработан понятийный аппарат.

Для блокирования и устранения указанных рисков в соответствии с Указом Президента Российской Федерации «О национальных целях и стратегических задачах развития Российской Федерации», при реализации национальной программы «Цифровая экономика Российской Федерации» Правительству Российской Федерации совместно с региональными органами государственной власти поручено к 2024 году обеспечить создание устойчивой и безопасной информационно-коммуникационной инфраструктуры высокоскоростной передачи, обработки и хранения больших объемов данных.

Библиографический список

1. Введение в теорию цифровой экономики / Центр социально-экономического прогнозирования им. Д.И. Менделеева, Сретенский клуб им. С.П. Курдюмова; А.В. Щербаков (отв. ред.), В.Г. Буданов, Л.А. Колесова, В.С. Курдюмов, А.В. Олескин. М.: Грифон, 2018. 104 с.

2. Маркова В.Д. Цифровая экономика: учебник / В.Д. Маркова. М.: ИНФРА-М, 2019, 186 с.

3. Что такое цифровая экономика? Тренды, компетенции, измерение [Текст]: докл. к XX Апр. междунар. науч. конф. по проблемам развития экономики и общества, Москва, 9-12 апр. 2019 г. / Г.И. Абдрахманова, К.О. Вишневский, Л.М. Гохберг и др.; науч. ред. Л.М. Гохберг; Нац. исслед. ун-т «Высшая школа экономики». М.: Изд. дом Высшей школы экономики, 2019. 82 с.

4. В потоке: мировая промышленность учится работать с большими данными. [Электронный ресурс] URL: http://digital-russia.rbc.ru/article-page_6.html (дата обращения 01.11.2020)

5. Что такое Big data. [Электронный ресурс] URL: <https://rb.ru/howto/chto-takoe-big-data/> (дата обращения 01.11.2020)

6. Что такое Big Data: характеристики, классификация, примеры. [Электронный ресурс] URL: <https://neurohive.io/ru/osnovy-data-science/big-data/> (дата обращения 01.11.2020)

7. Big Data – что такое системы больших данных? Развитие технологий Big Data. [Электронный ресурс] URL: <https://promdevelop.ru/big-data/> (дата обращения 01.11.2020)

Рекомендовано кафедрой
экономики и организации
производства БГТУ