

Технологии информационного общества

ИНДУСТРИЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Малахов Вадим Александрович

Кандидат исторических наук

Институт истории естествознания и техники им. С. И. Вавилова РАН, Центр истории организации науки и науковедения, старший научный сотрудник

Москва, Российская Федерация

yasonbh@mail.ru

Хохлов Юрий Евгеньевич

Кандидат физико-математических наук, доцент

Институт развития информационного общества, председатель совета директоров

РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО

Москва, Российская Федерация

yuri.hohlov@iis.ru

Шапошник Сергей Борисович

Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития, старший научный сотрудник

Петрозаводск, Российская Федерация

sergei.shaposhnik@gmail.com

Аннотация

В статье предложен подход, концептуальная схема и показатели для мониторинга индустрии работы с большими данными, под которой понимается совокупность компаний, производящих технологии, решения и услуги в области работы с большими данными. Концептуальная схема мониторинга основана на ролевой модели работы с большими данными, представленной в стандарте эталонной архитектуры больших данных Международной организации по стандартизации (ISO). В систему мониторинга включены 9 показателей, характеризующих масштабы, конкурентоспособность и потенциал индустрии работы с большими данными. Измерение показателя, связанного с числом стартапов в области индустрии работы с большими данными и основанного на базе данных Crunchbase, показало существенное отставание России по этому показателю от развитых стран (0,52% от всех стартапов в этой области).

Ключевые слова

большие данные; технологии работы с большими данными; индустрия работы с большими данными; мониторинг

Введение

Быстрый рост объемов данных, доступных для использования, а также развитие методов извлечения полезной информации из этих массивов данных приводят к росту спроса на технологии работы с большими данными в различных сферах деятельности. В ответ на этот спрос формируется новый сектор экономики – индустрия работы с большими данными, который объединяет поставщиков технологий, решений и услуг работы с большими данными. Развитие в России собственной индустрии работы с большими данными важно с нескольких точек зрения, включая решение задач создания новых высокопроизводительных рабочих мест, импортозамещения и обеспечения информационной безопасности при использовании больших данных в чувствительных сферах (оборона, национальная безопасность и др.).

© Малахов В.А., Хохлов Ю.Е., Шапошник С.Б., 2021.

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

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

https://doi.org/10.52605/16059921_2021_04_278

В данной статье предлагается комплексная методология мониторинга и оценки уровня отечественного производства технологий работы с большими данными на основе системы показателей, позволяющей измерять текущие масштабы индустрии работы с большими данными, ее глобальную конкурентоспособность и потенциал развития.

1 Определение предметной области мониторинга и оценки

Предметом мониторинга является состояние и потенциал развития сектора отечественной экономики, производящего технологии, решения и услуги, связанные с большими данными (далее – индустрии работы с большими данными).

С точки зрения классификации видов экономической деятельности (в России это ОКВЭД2) индустрия работы с большими данными является частью цифрового сектора экономики, под которым мы будем понимать группировку видов экономической деятельности, входящих в ИКТ-сектор и сектор контента и СМИ [1,2].

Наиболее удобным подходом к мониторингу индустрии работы с большими данными было бы формирование группировки кодов экономической деятельности, связанных с работой с большими данными, что позволило бы получать для этого из официальной статистики значения показателей, характеризующих масштабы, инновационную активность и т.д. Подобным образом, например, осуществляется мониторинг развития цифрового сектора в целом: доступные показатели, собираемые в рамках федерального статистического наблюдения, берутся в разрезе кодов ОКВЭД2, входящих в цифровой сектор экономики, и агрегируются для характеристики сектора. К сожалению, использование аналогичного подхода для такого специфического сектора экономики как производство технологий, решений и услуг, связанных с большими данными невозможно – коды классификатора видов экономической деятельности слишком «грубы» и широки, чтобы отразить нужные виды деятельности. В качестве примера можно привести тот факт, что для разработки программного обеспечения в ОКВЭД2 есть только один код (62.01 Разработка компьютерного программного обеспечения), который охватывает все виды программного обеспечения и не позволяет, например, выделить из них разработку технологий или приложений для работы с большими данными.

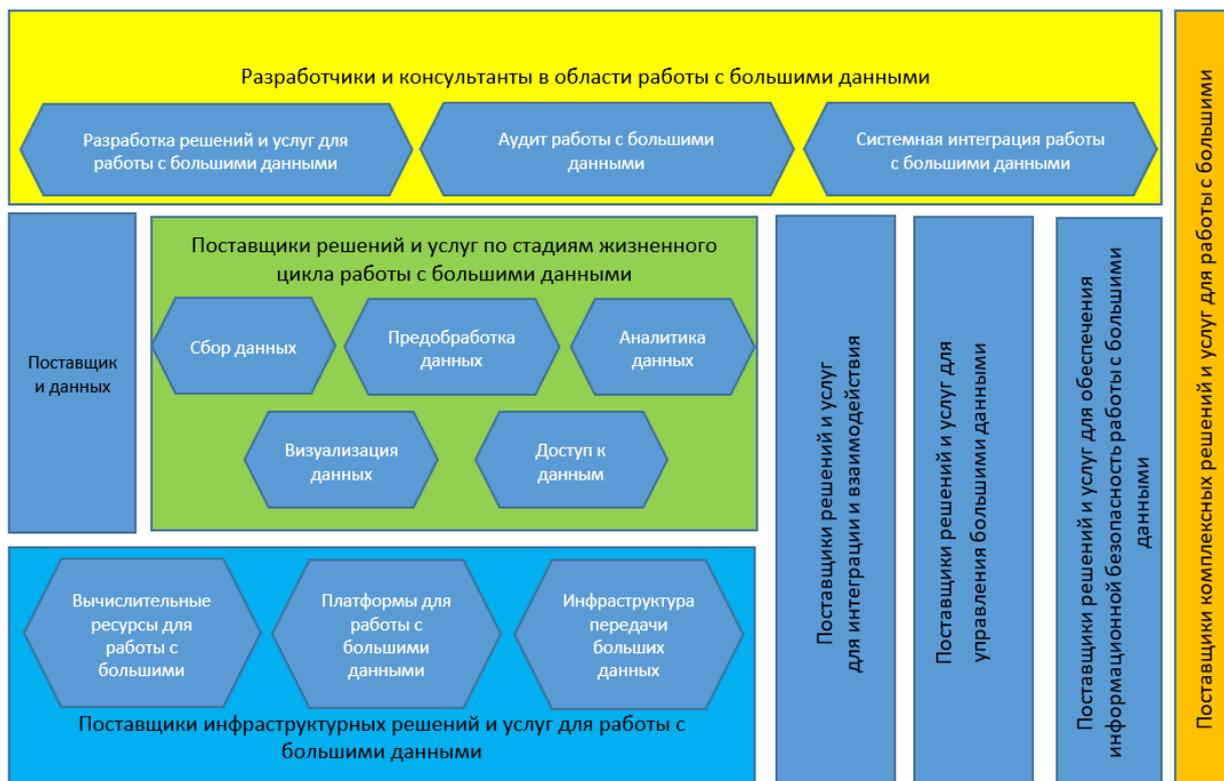


Рисунок 1. Производители (поставщики) технологий, решений и услуг для работы с большими данными

В этой ситуации для выделения индустрии работы с большими данными как сектора экономики был разработан подход, основанный на стандартах Международной организации по стандартизации (ISO), связанных с технологиями работы с большими данными, в частности, учитывающий стадии жизненного цикла работы с большими данными, зафиксированного в стандартах эталонной архитектуры больших данных [3]. В основу классификатора видов деятельности, которые предлагается включить в индустрию работы с большими данными, положена ролевая модель оперирования большими данными по жизненному циклу, а также обеспечивающие его вспомогательные и интеграционные виды деятельности. В рамках этого подхода к компаниям «индустрии работы с большими данными» предлагается относить поставщиков технологий, решений и услуг, схематично представленных на рисунке 1.

Указанные виды деятельности попадают в более широкие по охвату коды Общероссийского классификатора видов экономической деятельности (ОКВЭД2), представленные ниже в таблице 1.

Таблица 1. Соотношение классификации компаний индустрии работы с большими данными и кодов ОКВЭД2

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<i>Разработчики и консультанты в области работы с большими данными</i>	
<p>Разработка решений и технологий оказания услуг для работы с большими данными Включает формирование требований, планирование (концептуализацию), проектирование, разработку, тестирование технологий работы с большими данными и, при необходимости, передачу другому лицу технологической документации и необходимых знаний Включает формирование требований, планирование создания, проектирование, разработку, тестирование и подготовку к вводу в эксплуатацию систем для работы с большими данными и/или сервисов на их основе</p>	<p>62.01 Разработка компьютерного программного обеспечения 62.02.1 Деятельность по планированию, проектированию компьютерных систем 62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>
<p>Аудит массивов больших данных, решений и услуг для работы с большими данными Включает проведение аудита процессов и систем для работы с большими данными, связанного с проверкой достоверности источников и качества данных, операций с массивами больших данных, выполнением требований безопасности и защиты персональных данных, а также проверке соответствия заданному набору критериев аудита</p>	<p>62.02.2 Деятельность по обследованию и экспертизе компьютерных систем 63.1 Деятельность по обработке данных, предоставление услуг по размещению информации, деятельность порталов в информационно-коммуникационной сети Интернет</p>
<p>Системная интеграция работы с большими данными Включает формирование требований к решениям и услугам для работы с большими данными, включая требования к политике, системе (стратегического) управления, архитектуре, ресурсам, безопасности, качеству данных и требования бизнеса, деятельность по практической реализации разработанных требований, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>62.01 Разработка компьютерного программного обеспечения 62.02.1 Деятельность по планированию, проектированию компьютерных систем 62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<i>Поставщики больших данных</i>	
<p>Поставка данных</p> <p>Включает передачу, в том числе путем предоставления доступа, другому лицу массивов больших данных и прав на их использование, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>63.1 Деятельность по обработке данных, предоставление услуг по размещению информации, деятельность порталов в информационно-коммуникационной сети Интернет</p> <p>63.9 Деятельность в области информационных услуг прочая</p> <p>63.99 Деятельность информационных служб прочая, не включенная в другие группировки</p>
<i>Поставщики решений и услуг по стадиям жизненного цикла работы с большими данными</i>	
<p>Сбор (комплектование) данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для сбора данных и/или сервисов на их основе, связанных с выбором источников данных, приемом массивов больших данных, их регистрацией и буферизацией для последующих действий, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<p>Предобработка данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для предварительной обработки данных и/или сервисов на их основе, связанных с подготовкой полученных массивов больших данных к действиям над ними (управлению, преобразованию, обработке, анализу, визуализации, хранению, удалению и иным действиям), а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>58.29 Издание прочих программных продуктов</p> <p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.02.1 Деятельность по планированию, проектированию компьютерных систем</p>
<p>Аналитика данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для аналитики данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>62.02.4 Деятельность по подготовке компьютерных систем к эксплуатации</p>
<p>Визуализация данных</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для визуализации данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<p>Доступ к данным</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для управления доступом к массивам больших данных и/или сервисов на их основе, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<i>Поставщики инфраструктурных решений и услуг для работы с большими данными</i>	
<p>Предоставление платформенных решений для работы с большими данными</p> <p>Включает предоставление в пользование (аренду), или передачу другому лицу прав на использование платформенных решений для работы с большими данными и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p> <p>26.30 Производство коммуникационного оборудования</p>
<p>Предоставление вычислительных ресурсов для работы с большими данными</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для работы с большими данными и/или сервисов на их основе, связанных с вычислительными средствами (например, компьютерами, хранилищами данных) и/или физической средой (компьютерными залами, стойко-местами, электропитанием, кондиционерами и т.д.), обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.80 Производство незаписанных магнитных и оптических технических носителей информации</p> <p>61.10.3 Деятельность по предоставлению услуг по передаче данных и услуг доступа к информационно-коммуникационной сети Интернет</p>
<p>Предоставление телекоммуникационных решений для работы с большими данным</p> <p>Включает предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для работы с большими данными и/или сервисов на их основе, связанных с телекоммуникационными средствами для передачи массивов больших данных, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.03 Деятельность по управлению компьютерным оборудованием</p> <p>63.11 Деятельность по обработке данных, предоставление услуг по размещению информации и связанная с этим деятельность</p> <p>63.12 Деятельность веб-порталов</p>
<i>Поставщики решений и услуг, обеспечивающих работу с большими данными</i>	
<p>Поставка решений и услуг для интеграции и взаимодействия систем для работы с большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для интеграции и взаимодействия систем оперирования большими данными и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p>

Вид деятельности компаний индустрии работы с большими данными	Коды ОКВЭД2, которые могут иметь компании
<p>Поставка решений и услуг для управления большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для управления массивами больших данных и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	<p>26.30 Производство коммуникационного оборудования</p> <p>62.01 Разработка компьютерного программного обеспечения</p> <p>62.03 Деятельность по управлению компьютерным оборудованием</p>
<p>Поставка решений и услуг для обеспечения информационной безопасности работы с большими данными</p> <p>Включает поставку, предоставление в пользование (аренду) или передачу другому лицу прав на использование решений для обеспечения информационной безопасности систем оперирования массивами больших данных и/или сервисов на их основе, обеспечение функционирования, а также, при необходимости, передачу технологической документации и необходимых знаний</p>	
<p><i>Поставщики комплексных решений и услуг для работы с большими данными</i></p>	
<p>Поставка комплексных решений и услуг для работы с большими данными</p> <p>Включает несколько перечисленных выше одновременно выполняемых видов деятельности по работе с большими данными</p>	<p>61.10.3 Деятельность по предоставлению услуг по передаче данных и услуг доступа к информационно-коммуникационной сети Интернет</p> <p>62 Разработка компьютерного программного обеспечения, консультационные услуги в данной области и другие сопутствующие услуги</p> <p>63 Деятельность в области информационных технологий</p> <p>58.29 Издание прочих программных продуктов</p> <p>26.1 Производство элементов электронной аппаратуры и печатных схем (плат)</p> <p>26.20 Производство компьютеров и периферийного оборудования</p> <p>26.30 Производство коммуникационного оборудования</p> <p>26.80 Производство незаписанных магнитных и оптических технических носителей информации</p>

Таким образом, операционализируя данное выше определение индустрии работы с большими данными как сектора экономики, можно сказать, что это совокупность компаний, деятельность которых описывается кодами ОКВЭД2, входящими в цифровой сектор экономики (таблица 1) и выделяемыми в соответствии с видами деятельности, указанными на рисунке 1.

Аналогичный подход был реализован в проекте Европейской комиссии «Европейский рынок данных», в котором решалась задача выделить сферу деятельности компаний, оперирующих данными (data companies – компании, «основной деятельностью которых является производство и поставка продуктов, услуг и технологий, связанных с цифровыми данными») [4]. Для этого были определены более широкие коды классификатора видов экономической деятельности, в рамках которых выделялась совокупность компаний, оперирующих данными – в рамках проекта такими видами экономической деятельности были классы J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая) [4].

Выделение индустрии работы с большими данными и измерение показателей, ее характеризующих, проводится на базе выборочного представительного опроса организаций цифрового сектора экономики. В качестве критериев выделения компаний индустрии работы с большими данными и измерения показателей предлагается использовать предложенную в таблице 1 классификацию видов деятельности, составляющих индустрию работы с большими данными. Кроме того, в проводимом исследовании будет учитываться глобальное руководство и таксономия расходов на большие данные и аналитику, которое разработано и поддерживается в актуальном состоянии компанией IDC, многие годы занимающейся измерением глобального рынка аналитики больших данных [5]. Таксономия IDC технологий, решений и услуг работы с большими данными позволит выделить производителей и объемы выручки по продуктовым позициям в сфере работы с большими данными и проводить международные сопоставления. Наконец, для идентификации компаний индустрии работы с большими данными будет использоваться Единый реестр российских программ для электронных вычислительных машин и баз данных [6], для которого разработан более детальный классификатор, выделяющий программное обеспечение для работы с большими данными (см. в частности, раздел «10. Средства обработки и визуализации данных» и «11. Средства анализа данных»).

2 Обзор литературы

2.1. Обзор подходов, концептуальных схем и показателей, разработанных международными организациями и аналитическими компаниями для мониторинга сферы производства и рынка цифровых технологий

Как уже отмечалось, действующие классификаторы видов экономической деятельности, товаров и услуг не позволяют статистически выделить индустрию работы с большими данными и измерять на этой основе ее показатели.

Вместе с тем, международными организациями проделана большая работа по разработке и развитию стандартов измерения более широкого сектора экономики, частью которого является индустрия работы с большими данными, – информационной индустрии (цифрового сектора экономики, включающего ИКТ-сектор и сектор контента и СМИ). Опыт этого измерения, прежде всего в части предложенных показателей, использован в рамках данной работы.

Впервые определение и группировку видов экономической деятельности, относящихся к ИКТ-сектору, предложила ОЭСР в 1998 г. Впоследствии группировка дорабатывалась и адаптировалась к новой версии международного классификатора видов экономической деятельности, к ней была добавлена группировка сектора контента и СМИ [1]. В 2015 г. приказом Минкомсвязи России [2] были утверждены аналогичные группировки на основе ОКВЭД-2, согласованные с группировками ОЭСР.

Сектор ИКТ в рамках этого подхода определяется как совокупность видов экономической деятельности, связанных с производством товаров и услуг, предназначенных для выполнения функции обработки информации и коммуникации с использованием электронных средств, включая передачу и отображение информации. На основе этого определения был составлен перечень соответствующих кодов международного классификатора экономической деятельности. В состав сектора ИКТ вошли производство и оптовая торговля ИКТ-товарами (в частности,

компьютерного и коммуникационного оборудования), а также оказание ИКТ-услуг (телекоммуникационных услуг и услуг на основе информационных технологий – ИТ-услуг).

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

Также для измерения продукции цифрового сектора, ее экспорта и импорта ОЭСР были разработаны классификаторы товаров (первый вариант в 2003 г.) и услуг (2006 г.) ИКТ. Аналогичная работа была проделана для сектора контента и СМИ. Соответствующие классификаторы представляют собой собирательные группировки общих классификаторов товаров и услуг [1]. В 2015 г. приказом Минкомсвязи России были утверждены аналогичные группировки [2].

Для измерения цифрового сектора международные организации предложили и используют целый ряд показателей. Их набор ни одна из них не организует в виде концептуальной схемы, но их можно сгруппировать в несколько категорий: показатели масштабов, конкурентоспособности и потенциала развития информационной индустрии.

Сама ОЭСР для измерения ИКТ-сектора и сектора контента и СМИ использует несколько показателей. Два из них характеризуют размер (масштабы) сектора – (1) доля сектора в ВВП (или в добавленной стоимости, созданной в предпринимательском секторе); (2) доля занятых в секторе от всех занятых в экономике или в предпринимательском секторе [7, 8]. Международная торговля ИКТ-товарами и услугами представлена показателями, которые можно использовать для построения показателей конкурентоспособности – доли ИКТ-продукции в общем экспорте товаров и услуг [8,9].

Аналогичные показатели для ИКТ-сектора рекомендует международное Партнерство по измерению ИКТ для развития, добавляя в своем перечне ключевых показателей ИКТ к показателям международной торговли импорт товаров ИКТ в процентах от совокупного объема импорта [10].

Используемый Евростатом набор показателей ИКТ-сектора несколько шире. Помимо показателей размера сектора используется показатель темпов роста сектора, а также два показателя, характеризующих инновационный потенциал сектора ИКТ – (1) доля расходов на исследования и разработки сектора ИКТ в общих расходах на исследования и разработки и (2) доля персонала, занятого НИР в секторе от всего исследовательского персонала страны [11]. К характеристике потенциала развития ИКТ-сектора можно отнести показатель числа молодых быстрорастущих предприятий сектора (gazelles) [12].

В России НИУ ВШЭ совместно с Росстатом и Минцифры России с 2009 г. выпускает ежегодный статистический сборник «Индикаторы цифровой экономики» (до 2016 г. под названием «Индикаторы информационного общества»), посвященный мониторингу производства и использования цифровых технологий [13]. Один из разделов сборника посвящен мониторингу предприятий ИКТ-сектора и сектора контента и СМИ. Среди показателей – количество предприятий ИКТ-сектора, численность занятых, валовая добавленная стоимость и инвестиции в основной капитал. На основе выборочного обследования предприятий в сборнике также оценивается деловая активность организаций, оказывающих услуги в сфере ИКТ. Под деловой активностью понимаются конкурентоспособность, экономическое положение организации, объем продаж разработанного ПО на внутреннем рынке и за рубежом, спрос на услуги организации и т.д. Отдельный блок показателей характеризует структуру затрат на инновационную деятельность организаций ИКТ-сектора.

Возвращаясь к проблеме выделения предприятий индустрии работы с большими данными из предприятий цифрового сектора экономики необходимо обратить внимание на решение аналогичной проблемы, которое было предложено в рамках уже упоминавшегося выше проекта Европейской комиссии «Европейский рынок данных» [4]. Помимо прочего перечень разработанных в рамках проекта показателей включал в себя число компаний, оперирующих данными, и их выручку. Под компаниями, оперирующими данными, в исследовании понимались организации (стартапы, малые, средние предприятия или подразделения крупных компаний) которые занимаются разработкой и внедрением новых технологий и решений для работы с большими данными или усовершенствованием и обновлением старых. Таким образом, согласно приведенному выше определению, индустрия работы с большими данными включает в себя как «чистых игроков», чьим основным бизнесом являются продукты и услуги для работы с большими

данными, так и «смешанных игроков», которые предлагают как традиционные, так и основанные на работе с большими данными решения.

Оценка числа компаний, оперирующих данными и их выручки производилась в проекте Европейской комиссии с помощью выборочного обследования компаний, чья экономическая деятельность относится к классам J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая) согласно статистической классификации экономической деятельности в Европейском сообществе NACE Rev. 2 [14]. Выборка компаний состояла примерно из 1500 организаций. В нее вошли две группы компаний (по численности сотрудников): менее 250 сотрудников и 250 и более сотрудников. При этом опрос проводился среди независимых компаний, а не местных филиалов крупных транснациональных корпораций. Выборочное обследование компаний индустрии работы с большими данными было дополнено проведенным отдельно обследованием стартапов в данной отрасли. На основе полученных данных делалась оценка числа и выручки компаний, работающих в индустрии больших данных, в странах ЕС (доля компаний, производящих и занимающихся внедрением новых технологий, в выборке экстраполировалась на все компании, чья деятельность относится к классам J или M). Число компаний индустрии работы с большими данными в тех странах ЕС, где выборочного обследования не проводилось, также оценивалось по специальной методике с помощью экстраполяции данных, полученных в других странах. По аналогичной методике оценивалась выручка европейских компаний данного сектора экономики.

Таким образом, сегодня уже существуют достаточно проработанные подходы и наборы показателей, которые успешно применяются для мониторинга сферы производства и рынка цифровых технологий в России и мире. Данные показатели можно адаптировать для концептуальной схемы мониторинга сектора экономики, производящего технологии, решения и услуги работы с большими данными.

2.2. Обзор исследований и аналитических докладов, посвященных анализу индустрии данных и индустрии работы с большими данными

Несмотря на бурное развитие и использование технологий работы с большими данными в различных сферах человеческой деятельности, научная литература пока представлена сравнительно небольшим числом работ, посвященных измерению, оценке и анализу сектора экономики, разрабатывающего данный класс цифровых технологий, решений и сервисов на их основе. Это связано в том числе с обозначенными выше проблемами статистического учета предприятий-производителей и поставщиков продуктов и услуг для работы с большими данными, когда существующие классификации экономической деятельности оказываются слишком «грубыми» для данной сферы деятельности.

Так хотя поиск по специально разработанному поисковому образу, включающему все основные термины, связанные с технологиями работы с большими данными в библиографических базах данных платформы Web of Science [15], позволяет найти 108067 публикаций за 2016-2020 гг., абсолютное большинство из них посвящены анализу новых технологий, возможностей их применения и влиянию, которое они могут оказать на те или иные сферы деятельности. Поэтому для поиска релевантных статей необходимо сузить поисковый образ до двух терминов «индустрия работы с большими данными» (“big data industry”) и «компании, оперирующие большими данными» (“big data compan*”). Поиск проводился по полю «Тема», включающему названия, ключевые слова и аннотации статей. В результате по данному поисковому запросу было найдено 84 публикации, изданные в 2016-2020 гг. Далее содержание найденных работ было изучено с целью отбора релевантных статей. В общей сложности было отобрано 10 наиболее релевантных публикаций, посвященных анализу индустрии работы с большими данными, остальные 74 статьи были исключены из анализа, поскольку они не имели прямого отношения к исследуемой проблеме.

Значительная часть проанализированной научной литературы, посвященной индустрии работы с большими данными, носит описательный характер: в статьях рассматриваются основные технологии и решения для работы с большими данными существующие сегодня на рынке, их влияние на рынок и особенности ценообразования [16; 17]; влияние на индустрию больших данных отдельных технологий [18] однако не делается попытки измерить развитие данного сектора экономики. Некоторые исследователи делали попытки выявить и оценить факторы, влияющие на развитие индустрии работы с большими данными в отдельных регионах [19], а также взаимосвязь с «зеленой» экономикой [20].

Несколько иной подход выбрали авторы статьи [21], попробовав оценить развитие индустрии работы с большими данными в Китае по пяти категориям: 1) промышленный масштаб и уровень использования технологий работы с большими данными; 2) развитие предприятий-разработчиков и поставщиков технологий работы с большими данными; 3) технический уровень решений в области работы с большими данными; 4) особенности применения технологий работы с большими данными в промышленности Китая; 5) политическая среда развития индустрии работы с большими данными в стране. На основе анализа по перечисленным предметным областям авторы выявляют основные барьеры развития индустрии работы с большими данными и формируют рекомендации по их преодолению.

Схожее исследование по выявлению препятствий для развития индустрии работы с большими данными и разработке рекомендаций для корректировки политики в данной области было проведено в Корее [22]. Большинство подобных исследований строилось на обследовании определенной выборки предприятий-разработчиков и поставщиков технологий, услуг и решений работы с большими данными (как правило, это были самые крупные игроки на рынке), а выводы, сделанные на основе обследования данных выборок, распространялись на весь рынок [23; 24]. Из-за отсутствия надежных количественных статистических данных большинство исследований посвященных изучению индустрии работы с большими данными опираются на качественные методы, такие как интервью экспертов и представителей компаний, фокус-группы, SWOT-анализ и т.д. Исходя из результатов интервью и опросов, исследователи пытаются выявить основные направления развития индустрии работы с большими данными и разработать стратегии развития данного сектора экономики [25].

Таким образом, сегодня в мире не существует надежного методологического инструментария для оценки уровня развития сектора экономики, производящего технологии, решения и услуги, связанные с большими данными, а существующая статистика и действующие классификаторы не позволяют «вычленивать» индустрию работы с большими данными из общего сектора ИКТ. Поэтому большинство исследований индустрии работы с большими данными, посвящены либо отдельным предприятиям и технологиям (case-study), либо основаны на выборочных обследованиях сравнительно небольшого количества предприятий и экспертных опросах.

В описанных выше рекомендациях международных организаций и практике статистического наблюдения за развитием более широкого цифрового сектора экономики (или его части – ИКТ-сектора) сложились подход и система показателей, которые, как отмечалось выше, можно сгруппировать в три предметные области мониторинга: масштаб (доля сектора в ВВП и рабочей силы в занятом населении), конкурентоспособность (доля в общем экспорте и импорте, соотношение экспорта и импорта) и инновационная активность предприятий сектора, которая связана с потенциалом его развития (расходы предприятий на НИОКР и др.) [7-13].

3 Концептуальная схема и показатели мониторинга

3.1 Концептуальная схема предметной области

Для мониторинга индустрии работы с большими данными предложенная в предыдущем разделе группировка показателей была взята за основу концептуальной схемы, учитывая ее комплексный характер и многолетний опыт использования входящих в эту группировку показателей для измерения цифрового сектора. Вместе с тем состав показателей был доработан с учетом других систем мониторинга, потребностей аудитории мониторинга и дополнительных возможностей, которые предоставляет использование в рамках мониторинга развития и использования технологий работы с большими данными целого спектра методов измерения (в частности, экспертных оценок и опроса предприятий).

В таблице 2 представлена концептуальная схема и набор показателей для мониторинга индустрии работы с большими данными, которые будут рассчитываться для предприятий, выделяемых в соответствии с подходом и классификацией, представленными в предыдущем разделе.

Таблица 2. Концептуальная схема мониторинга производства технологий работы с большими данными

Масштабы производства технологий работы с большими данными	Конкурентоспособность производства технологий работы с большими данными	Потенциал производства технологий работы с большими данными
Доля индустрии работы с большими данными в общем объеме выручки цифрового сектора экономики	Доля российских поставщиков в общем объеме мирового экспорта технологий, товаров и услуг работы с большими данными	Доля российских стартапов в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными
Доля занятых в индустрии работы с большими данными, от общего числа занятых в цифровом секторе экономики	Доля российских производителей на мировом рынке технологий, товаров и услуг работы с большими данными	Соотношение отечественного и мирового уровней готовности технологий работы с большими данными
	Доля российских производителей на отечественном рынке технологий, товаров и услуг работы с большими данными	Доля проектов, связанных с производством технологий работы с большими данными, в общем числе проектов, поддержанных институтами развития
	Соотношение экспорта и импорта технологий, решений и услуг работы с большими данными	

3.2 Показатели масштабов производства

Как отмечалось в обзоре литературы, для характеристики масштабов отрасли экономики обычно используются несколько ключевых показателей, таких как доля этой отрасли в валовом внутреннем продукте, а также доля работников отрасли в общем числе занятого населения (международное Партнерство по измерению ИКТ для развития, добавляет к ним показатели международной торговли товарами ИКТ, такие как доля импорта ИКТ-товаров от общего объема импорта страны и доля экспорта ИКТ-товаров в общем объеме товарного экспорта страны [10]).

Для мониторинга индустрии работы с большими данными используется более широкая концептуальная схема, показатели международной торговли товарами и услугами работы с большими данными рассматриваются в контексте оценки конкурентоспособности индустрии. Что касается показателей, характеризующих масштаб индустрии работы с большими данными, предлагается их использовать в несколько модифицированном виде и рассчитывать долевые показатели добавленной стоимости и числа занятых не от показателей экономики в целом, а по отношению к цифровому сектору.

Выбор в качестве базы для расчета показателей цифрового сектора экономики, а не всей экономики (или ее предпринимательского сектора) связан с несколькими соображениями. Сам объем индустрии работы с большими данными пока невелик, значение показателей будет составлять доли процента от ВВП и с такими значениями, во-первых, трудно оперировать, а во-вторых, в таком варианте значения показателей будут чувствительны к изменению общей структуры экономики (например, повышение цен на энергетические ресурсы повышает долю ТЭК в экономике за счет других отраслей). Рассмотрение индустрии работы с большими данными в контексте общих показателей цифрового сектора позволит оценить относительную значимость этой индустрии во всем комплексе экономической деятельности, связанной с цифровыми технологиями, при этом всегда возможен простой пересчет показателей индустрии по отношению

к общеэкономическим показателям. Такой подход, отметим, принят в упомянутом выше исследовании «Европейский рынок данных» [4] – для характеристики «индустрии данных» используется, помимо общего числа компаний, оперирующих данными, и объема их выручки, показатели доли этих компаний не по отношению к экономике в целом, а в общем числе и выручке компаний, входящих в классы J (Деятельность в области информации и связи) и M (Деятельность профессиональная, научная и техническая).

В итоге, для характеристики масштабов индустрии работы с большими данными в данной работе будут использоваться следующие показатели:

- (МИБД-01-01) Доля индустрии работы с большими данными в общем объеме выручки цифрового сектора экономики

Использование предложенного показателя позволяет оценить масштаб индустрии работы с большими данными относительно всего цифрового сектора, а также его динамику. При этом абсолютные объемы производства товаров и услуг для работы с большими данными (выраженные, например, в рублевом эквиваленте), так же, как и динамика этих значений от года к году, могут быть вычислены напрямую на основе абсолютных объемов производства товаров и услуг цифрового сектора экономики.

- (МИБД-01-02) Доля занятых в индустрии работы с большими данными, от общего числа занятых в цифровом секторе экономики

Предложенный показатель демонстрирует социально-экономическую значимость индустрии работы с большими данными относительно всего цифрового сектора экономики; абсолютное значение и динамика от года к году также могут быть вычислены.

Так как официальные статистические данные не выделяют производство товаров и услуг для работы с большими данными из цифрового производства, основным инструментом для оценки указанных показателей является специализированное исследование, включающее представительные опросы предприятий цифрового сектора (см. выше).

3.3 Показатели конкурентоспособности производства

Конкурентоспособность отечественного производства технологий работы с большими данными характеризуется их востребованностью на отечественном и мировом рынках. Оценка конкурентоспособности проводится на основе измерений соотношения экспорта и импорта, доли российских производителей на рынках технологий работы с большими данными. Для мониторинга конкурентоспособности используется четыре показателя.

- (КИБД-02-01) Доля российских поставщиков в общем объеме мирового экспорта технологий, решений и услуг работы с большими данными

Показатель демонстрирует «вес» России среди мировых экспортеров технологий, товаров и услуг работы с большими данными. Более высокое значение показателей говорит о более высокой конкурентоспособности отечественных производителей.

- (КИБД-02-02) Доля российских производителей на мировом рынке технологий, решений и услуг работы с большими данными

Показатель демонстрирует востребованность продукции российских производителей на мировом рынке, а динамика показателя год к году демонстрирует изменение (рост или снижение) конкурентоспособности российских производителей. Более высокое значение показателя говорит о более высокой конкурентоспособности российских производителей.

- (КИБД-02-03) Доля отечественных поставщиков на российском рынке технологий, решений и услуг работы с большими данными

Показатель демонстрирует востребованность российских производителей на отечественном рынке, конкурентоспособность на внутреннем рынке товаров и услуг индустрии работы с большими данными по сравнению с импортными. Динамика показателя год к году демонстрирует изменение (рост или снижение) конкурентоспособности российских производителей на внутреннем рынке, более высокое значение показателя говорит о более высокой конкурентоспособности российских поставщиков.

- (КИБД-02-04) Соотношение экспорта и импорта технологий, решений и услуг работы с большими данными

Показатель демонстрирует, является ли Россия нетто-экспортером или нетто-импортером технологий работы с большими данными. Более высокое значение показателя говорит о способности российской индустрии работы с большими данными удовлетворять потребности отечественной экономики в технологиях, решениях и услугах работы с большими данными.

Предложенная система показателей позволяет с разных сторон оценить востребованность, а через нее – конкурентоспособность отечественного производства технологий работы с большими данными. При этом, оценивается конкурентоспособность как на отечественном, так и на мировом рынке.

Инструментом оценки показателей КИБД-02-01 – КИБД-02-04 является специализированное исследование, включающее обследование предприятий цифрового сектора экономики.

3.4 Показатели потенциала производства

Оценивая состояние индустрии работы с большими данными, важно характеризовать и потенциал ее развития – наличие условий для роста и конкурентоспособности отечественных производителей технологий работы с большими данными. Потенциал развития любой высокотехнологичной отрасли экономики обычно связывают с развитием существующих и разработкой новых технологий. В случае индустрии работы с большими данными потенциал оценивается тремя показателями – количеством новых стартапов и уровнем готовности наиболее зрелых технологий работы с большими данными, разрабатываемых отечественными производителями, относительно аналогичных технологий мировых производителей, а также долей проектов, связанных с производством технологий работы с большими данными, в общем числе проектов, поддержанных институтами развития.

- (ПИБД-03-01) *Доля российских стартапов в общем числе стартапов, связанных с технологиями работы с большими данными.*

Количество новых стартапов (относительно мирового их числа) демонстрирует активность отечественных разработчиков на рынке больших данных, которая, соответственно, потенциально может привести к разворачиванию производства новых технологий и решений для работы с большими данными, востребованных на отечественном и мировом рынках, к формированию новых сегментов рынков. Большее значение показателя свидетельствует о более высоком потенциале отечественной индустрии работы с большими данными.

Исходные данные для оценки показателя числа стартапов берутся из международной базы данных Crunchbase [26].

- (ПИБД-03-02) *Соотношение отечественного и мирового уровня готовности технологий работы с большими данными*

Уровень готовности технологии (Technology Readiness Level, TRL) и уровень готовности производства (Manufacturing Readiness Level, MRL) [27] являются общепринятыми методами измерения зрелости технологий и их производства, применимыми и для технологий работы с большими данными. Показатель будет измеряться методом экспертных опросов. Предполагается, что перед оценкой этого показателя экспертами будет выделяться перечень наиболее перспективных и востребованных технологий работы с большими данными, по каждой из которых будет выполнено соответствующее сопоставление отечественного и мирового уровней. Таким образом будет обеспечена актуальность оценки и сопоставимость с мировым уровнем.

- (ПИБД-03-03) *Доля проектов, связанных с производством технологий работы с большими данными, в общем числе проектов ИКТ, поддержанных институтами развития*

Показатель поддержки проектов в сфере производства технологий работы с большими данными отечественными институтами развития, работающими с проектами, находящимися на высоком уровне технологической и производственной готовности, отражает не только активность отечественных институтов развития в поддержке проектов, связанных с большими данными, но и является важной характеристикой потенциала развития индустрии работы с большими данными, отражая наличие в стране достаточного числа технологических инициатив, развивающих перспективные технологии.

4 Методология и результаты

Полномасштабная реализация разработанной методологии предполагает проведение комплексного обследования рынка и производителей технологий, товаров и услуг работы с большими данными, которое выходит за рамки возможностей пилотной реализации. В данном разделе представлены те показатели, измерение которых было возможно методом кабинетных исследований.

4.1 Методы измерения

Одним из показателей потенциала производства технологий работы с большими данными в России является количество новых российских стартапов в данной области и их доля от стартапов созданных во всем мире. В качестве источника информации о количестве созданных стартапов в России и мире в нашем исследовании используется база данных Crunchbase [26] – одна из крупнейших в мире баз данных, содержащая деловую информацию о частных и государственных компаниях. Данные Crunchbase поступают из различных каналов – от пользователей базы, представителей компаний, общедоступных источников или различных партнеров по сбору данных (например, инвесторских компаний). Хотя Crunchbase не претендует на всеобъемлющий охват своих данных и в некоторых случаях информация по одним отраслям или регионам может быть менее полной, чем по другим, на данный момент большинством ученых-экономистов эта база признается в качестве наиболее репрезентативной из существующих и активно используется в исследованиях посвященных стартапам, развитию инновационного бизнеса и цифровой экономики, в том числе при проведении международных сопоставлений [28; 29]. Данные Crunchbase также успешно применялись при исследовании российских стартапов [30].

В научной литературе нет консенсуса относительно того, какие компании можно считать стартапами, а какие нет. Как правило под стартапами понимают молодые, нацеленные на рост компании. Разные исследователи считают, что стартапами можно считать компании возрастом до одного, трех, шести или даже десяти лет [31]. Для целей данного мониторинга стартапами будут считаться компании, связанные с технологиями, товарами и услугами работы с большими данными, не старше 3 лет (т.е., например, если говорить про 2020 г., то стартапами мы считаем компании, зарегистрированные с 2018 по 2020 г. включительно).

Для отбора российских компаний, связанных с технологиями работы с большими данными, был сформирован поисковый образ, включающий 25 отраслей (по классификации Crunchbase), непосредственно относящихся к разработке и внедрению технологий работы с большими данными (таблица 3). Стоит отметить, что в классификации Crunchbase есть как достаточно широкие технологические отрасли (например, «Искусственный интеллект»), так и более узкие, привязанные к конкретным видам технологий (например, «Обработка естественного языка»). Учитывая, что компании в базе данных Crunchbase могут относиться одновременно к нескольким отраслям (например, к доставке еды и большим данным), разработанный нами поисковый образ позволяет создать по формальным признакам максимально репрезентативную выборку компаний, занимающихся разработкой и внедрением технологий, услуг и решений для работы с большими данными в различных отраслях экономики.

Таблица 3. Список отраслей по классификации Crunchbase, использованный для поиска компаний, занимающихся разработкой и (или) внедрением технологий, решений и услуг для работы с большими данными

№ п/п	Название отрасли
1	Аналитика (Analytics)
2	Искусственный интеллект (Artificial Intelligence)
3	Автономные транспортные средства (Autonomous Vehicles)
4	Большие данные (Big Data)
5	Бизнес-аналитика (Business Intelligence)
6	Облачные сервисы для работы с данными (Cloud Data Services)
7	Компьютерное зрение (Computer Vision)

№ п/п	Название отрасли
8	Центр обработки данных (Data Center)
9	Автоматизация центра обработки данных (Data Center Automation)
10	Интеграция данных (Data Integration)
11	Интеллектуальный анализ данных (Data Mining)
12	Хранение данных (Data Storage)
13	Визуализация данных (Data Visualization)
14	Распознавание лиц (Facial Recognition)
15	Распознавание изображений (Image Recognition)
16	Интеллектуальные системы (Intelligent Systems)
17	Машинное обучение (Machine Learning)
18	Обработка естественного языка (Natural Language Processing)
19	Предиктивная аналитика (Predictive Analytics)
20	Поисковая система (Search Engine)
21	Семантический поиск (Semantic Search)
22	Распознавание речи (Speech Recognition)
23	Интеллектуальный анализ текстов (Text Analytics)
24	Виртуальный помощник (Virtual Assistant)
25	Визуальный поиск (Visual Search)

4.2 Результаты

По данным Crunchbase, на 2020 г. в мире существовало 9619 стартапов (т.е. компаний, созданных в 2018–2020 гг.), занимающихся разработкой и внедрением технологий, решений и услуг для работы с большими данными, из них российских компаний – 50 (0,52%). Интересной тенденцией последних трех лет является постепенное сокращение количества стартапов. Так, с 2017 г. число стартапов сократилось более чем на треть: с 16 тыс. до менее 10 тыс. во всем мире, в России – с 81 до 51. Особенно резкое сокращение наблюдается в 2020 г., что, по-видимому, связано с глобальной пандемией коронавируса, а также с тем, что данные базы Crunchbase могут пополняться с некоторым запозданием.

Абсолютным лидером в мире по количеству интересующих стартапов являются США. В то же время их доля в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными, в последние годы стабильно падает. Так, в 2017 г. в Crunchbase была информация о 5610 молодых американских компаний, что составляло 37,4% от всех зарегистрированных в базе стартапов, работающих с большими данными. А уже в 2020 г. эта доля сократилась до 33% (3179 компаний). Еще более резкое падение числа новых компаний наблюдается в Китае: доля китайских компаний в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными упала с 6,1% (909 стартапов) в 2017 г. до 1,9% (181 стартап). Рост по данному показателю среди стран-лидеров в последние годы наблюдался только в Германии (2,6% до 3,2%) и Японии (с 1,8% до 2,3%). Данные о динамике доли стартапов стран, являющихся технологическими лидерами, в общем числе стартапов, связанных с технологиями, товарами и услугами работы с большими данными, за последние пять лет представлены на рисунке 1.

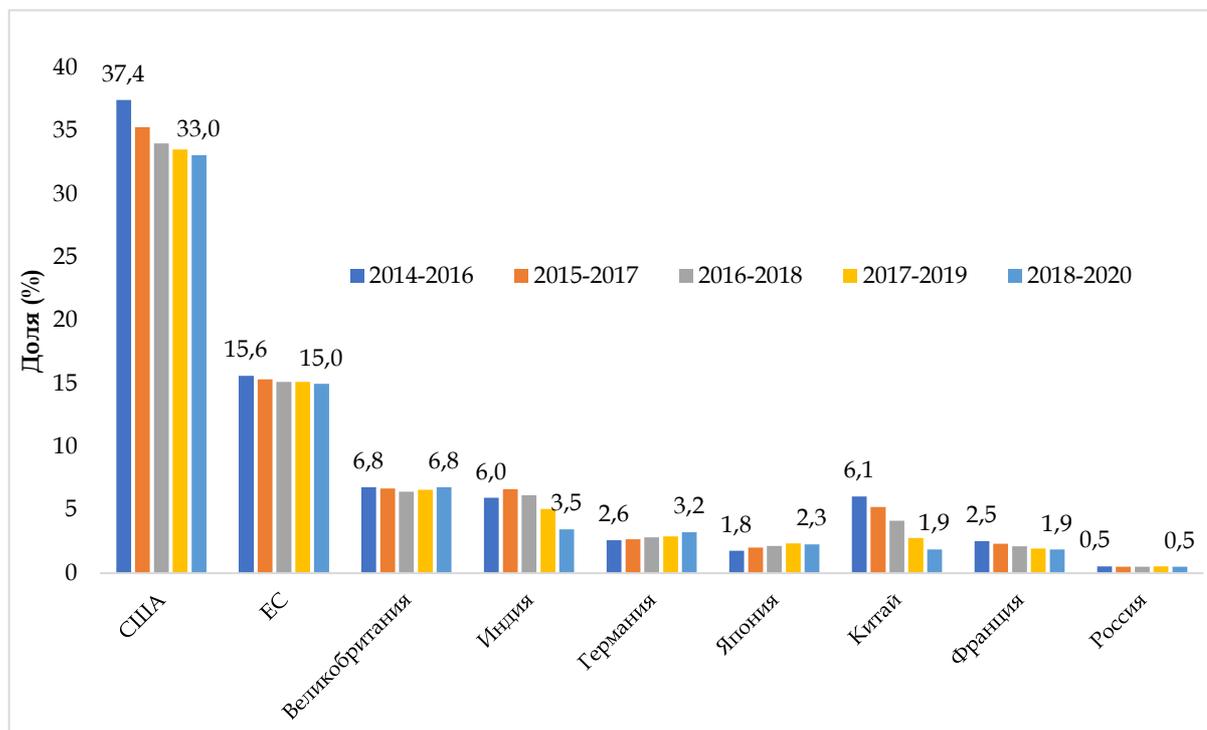


Рисунок 1. Динамика изменения доли новых стартапов различных стран в общем числе мировых стартапов, связанных с технологиями, товарами и услугами работы с большими данными (2014–2020 гг.)
Источник: составлено авторами на основе данных Crunchbase

Как видно из графика, Россия серьезно отстает от большинства развитых и многих развивающихся стран по значению показателя, характеризующего долю российских стартапов в общем числе мировых стартапов, связанных с технологиями, товарами и услугами работы с большими данными. Это свидетельствует о сравнительно низком уровне развития индустрии больших данных в России. В то же время, в отличие от большинства стран-лидеров в данной области, доля российских компаний среди мировых стартапов стабильно держится на уровне 0,5%, не снижаясь последние пять лет.

Заклучение

В условиях отсутствия статистических инструментов для измерения индустрии работы с большими данными (необходимой для этого детализации классификаторов видов экономической деятельности, товаров и услуг работы с большими данными) для мониторинга этого важного сектора отечественной экономики был разработан представленный в статье подход, включающий:

- определение и классификацию видов деятельности, которые относятся к индустрии работы с большими данными (основана на модели работы с большими данными, представленной в стандарте эталонной архитектуры больших данных Международной организации по стандартизации);
- концептуальную схему и систему показателей для мониторинга;
- методы измерения показателей, основанные на опросе предприятий цифрового сектора и потребителей технологий, решений и услуг работы с большими данными.

Одним из основных направлений дальнейшего развития методологии мониторинга является разработка таксономии технологий, товаров (решений) и услуг работы с большими данными, а также ее тестирование в ходе проведения опроса предприятий цифрового сектора и потребителей технологий, решений и услуг работы с большими данными.

Измерение одного из показателей методологии показало, что на такой авторитетной международной площадке, как Crunchbase российские компании и стартапы пока составляют небольшую долю. Одним из возможных факторов низкого значения показателя среди прочего может быть недостаточно полное освещение данных по российским компаниям в базе Crunchbase по сравнению с компаниями из Северной Америки или Западной Европы. Учитывая, что

источниками данных Crunchbase являются в том числе и представители компаний, для исправления ситуации можно рекомендовать в обязательном порядке всем российским стартапам вносить информацию в данную базу.

Разработанный подход к мониторингу индустрии работы с большими данными может быть использован для создания систем мониторинга развития, производства и использования «сквозных» цифровых технологий – для них возникают аналогичные проблемы при определении производителей решений и услуг, основанных на этих технологиях.

Благодарности

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

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

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

Литература

1. OECD. OECD Guide to Measuring the Information Society 2011. Paris: OECD Publishing, 2011. <https://doi.org/10.1787/9789264113541-en>
2. Приказ Минкомсвязи России от 7 декабря 2015 года № 515 «Об утверждении собирательных классификационных группировок "Сектор информационно-коммуникационных технологий" и "Сектор контента и средств массовой информации"». URL: <http://docs.cntd.ru/document/420327966> (дата обращения 01.06.2021).
3. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture. URL: <https://www.iso.org/standard/71277.html> (дата обращения 01.06.2021).
4. The European Data Market Study: Final Report. By IDC and Open Evidence. 2017. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirects=0&d=1 (дата обращения 01.06.2021)
5. Goepfert J. et al. IDC's Worldwide Big Data and Analytics Spending Guide Taxonomy, IDC. URL: <https://www.idc.com/getdoc.jsp?containerId=US47485920> (дата обращения 01.06.2021)
6. Единый реестр российских программ для электронных вычислительных машин и баз данных. URL: <https://reestr.digital.gov.ru/> (дата обращения 01.06.2021)
7. OECD. Size of the ICT Sector. 2013. URL: <https://www.oecd-ilibrary.org/docserver/factbook-2011-72-en.pdf?expires=1632472701&id=id&accname=guest&checksum=8C61EC5E182618032918749217A03978> (дата обращения 01.06.2021)
8. OECD. Measuring the Digital Transformation. A Roadmap for the Future. Paris; OECD Publishing, 2019. <https://doi.org/10.1787/9789264311992-en> (дата обращения 01.06.2021)
9. OECD Key ICT Indicators. URL: <https://www.oecd.org/sti/ieconomy/oecdkeyictindicators.htm> (дата обращения 01.06.2021)
10. Core list of indicators. March 2016 version. Partnership on Measuring ICT for Development. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf (дата обращения 01.06.2021).
11. ICT sector (isoc_se). Reference Metadata in Euro SDMX Metadata Structure (ESMS). Eurostat, the statistical office of the European Union. URL: https://ec.europa.eu/eurostat/cache/metadata/en/isoc_se_esms.htm (дата обращения 01.06.2021).

12. Eurostat. Browse statistics by theme. Digital economy and society. Database. ICT sector. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (дата обращения 01.06.2021)
13. Индикаторы цифровой экономики: 2020: статистический сборник / Г. И. Абдрахманова, К. О. Вишневецкий, Л. М. Гохберг и др.; Нац. исслед. ун-т «Высшая школа экономики». – М.: НИУ ВШЭ, 2020. URL: <https://issek.hse.ru/mirror/pubs/share/387609461.PDF> (дата обращения 01.06.2021).
14. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European. - Luxembourg: Office for Official Publications of the European Communities, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (дата обращения 01.06.2021)
15. Т.В. Ершова, Ю.Е. Хохлов, С.Б. Шапошник. Методология мониторинга развития и использования технологий работы с большими данными // Информационное общество. 2021. № 4–5. С. 2–32. https://doi.org/10.52605/16059921_2021_04_02
16. Rojas-Torres D., Kshetri N. Big data solutions for micro-, small-, and medium-sized enterprises in developing countries // IT Professional. 2019. Vol. 21. №. 5. P. 67-70. <https://doi.org/10.1109/MITP.2019.2932236>
17. Liang F. et al. A survey on big data market: Pricing, trading and protection // Ieee Access. 2018. Vol. 6. P. 15132–15154. <https://doi.org/10.1109/ACCESS.2018.2806881>
18. Ahmad M., Rashid T., Mishra D. K. Impact of wireless technology on future of big-data industry // 2014 Conference on IT in Business, Industry and Government (CSIBIG). IEEE, 2014. P. 1–4, <https://doi.org/10.1109/CSIBIG.2014.7056931>
19. Li C., Pan K. Research on financing efficiency of big data industry based on three stage DEA-Taking Guizhou Province as an example // Advances in Social Science, Education and Humanities Research (ASSEHR), 2018. Vol. 130. P. 439–442. <https://doi.org/10.2991/ifmeita-17.2018.74>
20. Zhang N., Zhang H., Zhang M. The Harmonious Development of Big Data Industry and Ecological Civilization Construction in Guizhou // 2018 6th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM 2018). Atlantis Press, 2019. P. 192–198. <https://doi.org/10.2991/ieesasm-18.2019.36>
21. Liu Y. et al. An overview of big data industry in China // China Communications. 2014. Vol. 11. № 12. P. 1–10. <https://doi.org/10.1109/CC.2014.7019834>
22. Kwon T. H., Kwak J. H., Kim K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea // Technological Forecasting and Social Change. – 2015. – Т. 96. – С. 144–152. <https://doi.org/10.1016/j.techfore.2015.03.017>
23. Caputo F. et al. Over the mask of innovation management in the world of Big Data // Journal of Business Research. 2020. Vol. 119. P. 330–338. <https://doi.org/10.1016/j.jbusres.2019.03.040>
24. Tang C. The data industry: The business and economics of information and big data. Hoboken: John Wiley & Sons, 2016, 195 p.
25. Weng W. H., Lin W. T. Development Trends and Strategy Planning in Big Data Industry // Contemporary Management Research. 2014. Vol. 10. № 3. P. 203–214. <https://doi.org/10.7903/cmr.12288>
26. Crunchbase. 2021. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirects=0&d=1 (дата обращения 01.06.2021).
27. ГОСТ Р 58048–2017 Трансфер технологий. Методические указания по оценке уровня зрелости технологий; ГОСТ Р 57194.1–2016 Трансфер технологий. Общие положения. URL: <https://docs.cntd.ru/document/1200158331> (дата обращения 01.06.2021).
28. Dalle J. M., Besten M., Menon C. Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers, No. 2017/08. Paris: OECD Publishing, 2017. 29 p. <https://doi.org/10.1787/6c418d60-en>
29. Kézai P. K., Fischer S., Lados M. Smart economy and startup enterprises in the Visegrád Countries – A comparative analysis based on the crunchbase database // Smart Cities. 2020. Vol. 3. №. 4. P. 1477–1494. <https://doi.org/10.3390/smartcities3040070>
30. Akhmetova J. A. The Impact of Qualitative Determinants on Startup Investments in Russia // Наука, технологии и бизнес: сборник материалов II межвузовской заочной конференции аспирантов, соискателей и молодых ученых, Москва, 28–29 апреля 2020 года. Москва:

- Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет), 2020. С. 35–41.
31. Иншаков М. О., Орлова А. А. Инновационные стартапы в России: проблемы создания и маркетингового продвижения // Вестник Волгоградского государственного университета. Серия 3: Экономика. Экология. 2014. №. 1. С. 66–76.

BIG DATA INDUSTRY

Malakhov, Vadim Aleksandrovich

Candidate of historical sciences

Vavilov Institute for the History of Science and Technology of the Russian Academy of Sciences, Center for the history of science organization and scientific research, senior researcher

Moscow, Russia

yasonbh@mail.ru

Hohlov, Yuri Eugenyevich

Candidate of physical and mathematical sciences, associate professor

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

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

Moscow, Russia

yuri.hohlov@iis.ru

Shaposhnik, Sergei Borisovich

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

Petrozavodsk, Russian Federation

sergei.shaposhnik@gmail.com

Abstract

The article proposes an approach, a conceptual framework and indicators for monitoring the big data industry, which is understood as a set of companies producing technologies, solutions and services in the field of big data. The conceptual framework for monitoring is based on the role model of big data, as presented in the International Organization for Standardization (ISO) standard for big data reference architecture. The monitoring framework includes 9 indicators characterizing the scale, competitiveness, and potential of the big data industry. Measurement of the indicator related to the number of startups in the big data industry, based on the Crunchbase database, showed a significant lag of Russia from developed countries (0.52% of all startups in this area).

Keywords

big data; big data technologies; big data industry; monitoring

References

1. OECD (2011). OECD Guide to Measuring the Information Society 2011, OECD Publishing, Paris, <https://doi.org/10.1787/9789264113541-en>
2. Order of the Ministry of Communications and Mass Media of Russia No. 515 of December 7, 2015 "On Approval of the Collective Classification Groups "Information and Communication Technologies Sector" and "Content and Mass Media Sector" URL: <http://docs.cntd.ru/document/420327966> (accessed: 01.06.2021).
3. ISO/IEC 20547-3:2020 Information technology – Big data reference architecture – Part 3: Reference architecture. URL: <https://www.iso.org/standard/71277.html> (accessed: 01.06.2021).
4. The European Data Market Study: Final Report. By IDC and Open Evidence. 2017. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirec%20ts=0&d=1 (accessed: 01.06.2021).
5. IDC's Worldwide Big Data and Analytics Spending Guide Taxonomy, IDC
6. Edinyj reestr rossijskih programm dlja jelektronnyh vychislitel'nyh mashin i baz dannyh. URL: <https://reestr.digital.gov.ru/> (accessed: 01.06.2021).
7. OECD (2013). SIZE OF THE ICT SECTOR. URL: <https://www.oecd-ilibrary.org/docserver/factbook-2011-72-en.pdf?expires=1632472701&id=id&acname=guest&checksum=8C61EC5E182618032918749217A03978> (accessed: 01.06.2021).

8. OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264311992-en>.
9. OECD Key ICT Indicators. URL: <https://www.oecd.org/sti/ieconomy/oecdkeyictindicators.htm>
10. Core list of indicators. March 2016 version. Partnership on Measuring ICT for Development. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf (accessed: 01.06.2021).
11. ICT sector (isoc_se). Reference Metadata in Euro SDMX Metadata Structure (ESMS). Eurostat, the statistical office of the European Union. URL: https://ec.europa.eu/eurostat/cache/metadata/en/isoc_se_esms.htm (accessed: 01.06.2021).
12. Eurostat. Browse statistics by theme. Digital economy and society. Database. ICT sector. URL: <https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> (accessed: 01.06.2021).
13. Indikatory cifrovoj jekonomiki: 2020: statisticheskij sbornik / G. I. Abdrahmanova, K. O. Vishnevskij, L. M. Gohberg i dr.; Nac. issled. un-t «Vysshaja shkola jekonomiki». – M.: NIU VShJe, 2020. URL: <https://issek.hse.ru/mirror/pubs/share/387609461.PDF> (accessed: 01.06.2021).
14. Eurostat. NACE Rev. 2. Statistical classification of economic activities in the European. - Luxembourg: Office for Official Publications of the European Communities, 2008. URL: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (accessed: 01.06.2021).
15. T.V. Ershova, Y.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // Informatsionnoye obshchestvo. 2021. № 4–5. S. 2–32. https://doi.org/10.52605/16059921_2021_04_02
16. Rojas-Torres D., Kshetri N. Big data solutions for micro-, small-, and medium-sized enterprises in developing countries // IT Professional. 2019. Vol. 21. №. 5. P. 67-70. <https://doi.org/10.1109/MITP.2019.2932236>
17. Liang F. et al. A survey on big data market: Pricing, trading and protection // Ieee Access. 2018. Vol. 6. P. 15132-15154. <https://doi.org/10.1109/ACCESS.2018.2806881>
18. Ahmad M., Rashid T., Mishra D. K. Impact of wireless technology on future of big-data industry // 2014 Conference on IT in Business, Industry and Government (CSIBIG). IEEE, 2014. P. 1-4, <https://doi.org/10.1109/CSIBIG.2014.7056931>
19. Li C., Pan K. Research on financing efficiency of big data industry based on three stage DEA-Taking Guizhou Province as an example // Advances in Social Science, Education and Humanities Research (ASSEHR), 2018. Vol. 130. P. 439-442. <https://doi.org/10.2991/ifmeita-17.2018.74>
20. Zhang N., Zhang H., Zhang M. The Harmonious Development of Big Data Industry and Ecological Civilization Construction in Guizhou // 2018 6th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM 2018). Atlantis Press, 2019. P. 192-198. <https://doi.org/10.2991/ieesasm-18.2019.36>
21. Liu Y. et al. An overview of big data industry in China // China Communications. 2014. Vol. 11. № 12. P. 1-10. <https://doi.org/10.1109/CC.2014.7019834>
22. Kwon T. H., Kwak J. H., Kim K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea // Technological Forecasting and Social Change. – 2015. – T. 96. – S. 144-152. <https://doi.org/10.1016/j.techfore.2015.03.017>
23. Caputo F. et al. Over the mask of innovation management in the world of Big Data // Journal of Business Research. 2020. Vol. 119. P. 330-338. <https://doi.org/10.1016/j.jbusres.2019.03.040>
24. Tang C. *The data industry: The business and economics of information and big data*. Hoboken: John Wiley & Sons, 2016, 195 p.
25. Weng W. H., Lin W. T. Development Trends and Strategy Planning in Big Data Industry // Contemporary Management Research. 2014. Vol. 10. № 3. P. 203-214. <https://doi.org/10.7903/cmr.12288>
26. Crunchbase. 2021. URL: https://sites.google.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attredirets=0&d=1 (accessed: 01.06.2021).

27. GOST R 58048-2017 Transfer tehnologij. Metodicheskie ukazaniya po ocenke urovnja zrelosti tehnologij; GOST R 57194.1-2016 Transfer tehnologij. Obshhie polozheniya. URL: <https://docs.cntd.ru/document/1200158331> (accessed: 01.06.2021).
28. Dalle J. M., Besten M., Menon C. Using Crunchbase for economic and managerial research. OECD Science, Technology and Industry Working Papers, No. 2017/08. Paris: OECD Publishing, 2017. 29 p. <https://doi.org/10.1787/6c418d60-en>
29. Kézai P. K., Fischer S., Lados M. Smart economy and startup enterprises in the Visegrád Countries – A comparative analysis based on the crunchbase database // Smart Cities. 2020. Vol. 3. №. 4. P. 1477-1494. <https://doi.org/10.3390/smartcities3040070>
30. Akhmetova J. A. The Impact of Qualitative Determinants on Startup Investments in Russia // Nauka, tehnologii i biznes: sbornik materialov II mezhvuzovskoj zaochnoj konferencii aspirantov, soiskatelej i molodyh uchenyh, Moskva, 28–29 aprelja 2020 goda. Moskva: Moskovskij gosudarstvennyj tehničeskij universitet imeni N. Je. Baumana (nacional'nyj issledovatel'skij universitet), 2020. P. 35-41.
31. Inshakov M. O., Orlova A. A. Innovacionnye startapy v Rossii: problemy sozdaniya i marketingovogo prodvizheniya // Vestnik Volgogradskogo gosudarstvennogo universiteta. Serija 3: Jekonomika. Jekologija. 2014. №. 1. P. 66-76