

Информационное общество: политика и факторы развития

## ЦИФРОВАЯ ИНФРАСТРУКТУРА ДЛЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

**Ершов Петр Сергеевич**

*Институт развития информационного общества, руководитель Дирекции региональных программ  
Москва, Российская Федерация  
peter.erшов@iis.ru*

**Хохлов Юрий Евгеньевич**

*Кандидат физико-математических наук, доцент  
Институт развития информационного общества, председатель Совета директоров  
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО  
Москва, Российская Федерация  
yuri.hohlov@iis.ru*

### Аннотация

*Разработаны концептуальная схема и набор показателей мониторинга и оценки цифровой инфраструктуры для работы с большими данными. Концептуальная схема состоит из трех компонентов: телекоммуникационная инфраструктура; вычислительная инфраструктура; платформенная инфраструктура. Приведены расчеты основных показателей для России и отдельных сфер деятельности, проведены международные сопоставления по состоянию на конец 2020 года. Отмечено существенное отставание России от стран ЕС по отдельным инфраструктурным показателям.*

### Ключевые слова

*цифровая технология; большие данные; цифровая инфраструктура; информационная инфраструктура; телекоммуникационная инфраструктура; сеть передачи данных; вычислительная инфраструктура; цифровая платформа; мониторинг и оценка; Big Data for Digital Economy, BD4DE*

### Введение

Наличие развитой цифровой инфраструктуры, удовлетворяющей растущим потребностям в передаче, хранении и обработке больших данных, является одним из ключевых факторов, влияющих на развитие цифровой экономики в современном мире. Основу инфраструктуры для работы с большими данными составляют инфраструктура передачи данных на основе телекоммуникационных фиксированных и мобильных сетей, вычислительная инфраструктура на основе центров обработки данных, а также экосистемы цифровых платформ, предоставляющих богатый спектр цифровых сервисов для участников многосторонних рынков.

Сегодняшние потребности в высокоскоростной инфраструктуре передачи данных возросли настолько быстро, что глобальный интернет испытывает серьезные трудности, связанные не только с пандемией. По данным компании TeleGeography [1] в 2020 году средний международный интернет-трафик увеличился на 48%, в то время как пропускная способность выросла на 35%, что является самым большим ростом с 2013 года. В период с 2016 по 2020 годы средний и пиковый международный интернет-трафик увеличивался в среднем на 30% в год, тогда как среднегодовой темп роста пропускной способности составлял 29%. Эта тенденция во многом обусловлена необходимостью передачи больших массивов данных, для чего и нужна соответствующая инфраструктура как на глобальном уровне, так и внутри страны.

Быстро накапливающиеся объемы данных требуют наличия все большего количества вычислительных ресурсов. По оценкам компании Global Industry Analysts [2] объем мирового

---

© Ершов П.С., Хохлов Ю.Е., 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\\_110](https://doi.org/10.52605/16059921_2021_04_110)

рынка услуг центров обработки данных будет расти начиная с 2020 года на 13,4%. Высокие темпы роста прогнозируются в Канаде (+11,8% в год), Германии (+10,6%) и Японии (+9,5%). При этом китайский рынок будет увеличиваться в объеме еще быстрее – на 17,5% ежегодно. Глобальными лидерами рынка услуг центров обработки данных по состоянию на май 2021 года являются транснациональные компании Amazon, Microsoft и Google, владеющие более 50% крупнейших центров обработки данных в мире. Страны-лидеры цифровой экономики активно инвестируют в развитие передовых ИКТ, особое место среди которых занимают высокопроизводительные вычисления (high performance computing), применяемые для параллельной обработки данных при решении сложных вычислительных задач. Прогнозируется, что данный сегмент рынка вычислительных ресурсов тоже будет иметь среднегодовой темп роста 5,5% [3].

Одним из ключевых драйверов цифровой трансформации экономики стали онлайн-цифровые платформы, появление которых принципиально изменило способы ведения бизнеса в отраслях [4]. Под цифровыми платформами, как правило, понимают реализованные с помощью цифровых технологий «многосторонние торговые площадки с бизнес-моделями, которые предоставляют производителям и потребителям возможность создавать обоюдовыгодные ценности, взаимодействуя друг с другом» [5]. Стремительный рост многосторонних цифровых платформ ведет к повышению производительности, снижению транзакционных издержек при взаимодействии, обеспечивает мгновенную и непрерывную связь между людьми, облегчает сотрудничество и кооперацию между организациями, стирая территориальные границы и преодолевая часовые пояса, а также усиливает конкурентоспособность национальных рынков [6], [7]. Цифровые платформы предоставляют новые возможности участникам рынка за счет взаимного влияния двух эффектов: сетевого и инновационного. Сетевой эффект возникает благодаря использованию цифровых технологий при осуществлении транзакций и объединяет миллионы потребителей на единой площадке. Инновационный эффект возникает там, где операторы платформ предоставляют в общее пользование данные и сервисы, способствуя тем самым совместным технологическим инновациям различных игроков рынка. В последние годы еще одним трендом в развитии стало формирование экосистем цифровых платформ, порождающих новые виды экономических отношений и еще больше усиливающих сетевые и инновационные эффекты за счет взаимной диффузии доселе отдельно функционирующих сегментов рынка [8].

Поэтому наличие цифровых платформ в стране или отдельной сфере деятельности является не менее важным фактором, влияющим на развитие цифровой экономики, чем наличие зрелой телекоммуникационной или вычислительной инфраструктуры, о которых шла речь выше.

В данной статье представлена концептуальная схема компонентов мониторинга развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации (Big Data for Digital Economy, BD4DE) [9], относящихся к цифровой инфраструктуре для работы с большими данными в масштабах страны и в отдельных сферах деятельности, а также результаты проведенной на основе разработанной методологии (и части предложенных показателей) оценки этого фактора по данным в 2020 года.

## **1 Определение предметной области мониторинга и оценки**

Предметом мониторинга в рамках данного исследования является состояние цифровой инфраструктуры, которое может как способствовать удовлетворению социальных и экономических потребностей основных заинтересованных сторон (граждан, бизнеса, власти, научно-образовательного сообщества) благодаря использованию цифровых технологий для работы с данными, так и выступать препятствием на пути развития информационного общества и цифровой экономики.

### **1.1 Обзор подходов к мониторингу и оценке цифровой инфраструктуры для работы с большими данными**

Наблюдение и оценка состояния ИКТ-инфраструктуры как необходимого условия для развития информационного общества имеет давнюю историю и ведется как на международном уровне, прежде всего, Международным союзом электросвязи (далее – МСЭ) [10], так и на национальном уровне соответствующими администрациями связи и/или национальными статистическими ведомствами. За эти годы сложился общепризнанный перечень показателей, характеризующих в первую очередь уровень развития как телекоммуникационной инфраструктуры передачи данных,

так и вычислительной инфраструктуры для хранения и обработки данных. Эти наблюдения ведутся такими международными организациями, как МСЭ [11], Организация экономического сотрудничества и развития (далее – ОЭСР) [12], Евростат [13]. В Российской Федерации долгосрочное наблюдение по аналогичным показателям ведется Росстатом [14].

Использование традиционных показателей ИКТ-инфраструктуры для оценки цифровой инфраструктуры передачи, хранения и обработки больших данных нашло широкое распространение и применяется как для международных сопоставлений, так и между различными сферами деятельности. В то же время следует отметить, что для инфраструктурного уровня, связанного с активно формирующимися в последние годы на глобальном, национальном или отраслевом уровнях экосистемами цифровых платформ [15], [7] еще не разработаны устоявшиеся инструменты мониторинга и системы показателей, что приводит к трудностям при оценке и проведении международных или межотраслевых сопоставлений. В данной работе будет предложен авторский подход к мониторингу цифровых платформ, развивающий направление, обозначенное в работе [16], как к необходимому условию успешного развития и использования технологий работы с большими данными при решении социальных или экономических задач.

С учетом вышесказанного предметная область мониторинга цифровой инфраструктуры работы с большими данными должна содержать три основных компонента, позволяющих оценить наличие и доступность развитой телекоммуникационной инфраструктуры, наличие и доступность вычислительных ресурсов для хранения и обработки больших данных, наличие и доступность цифровых платформ для развития и использования технологий работы с большими данными.

## 1.2 Обзор литературы по мониторингу и оценке цифровой инфраструктуры для работы с большими данными

Для разработки концептуальной схемы мониторинга цифровой инфраструктуры для работы с большими данными библиометрическим методом был проведен анализ релевантных научных публикаций, индексируемых в платформе Web of Science (WoS).

На первом этапе из библиографической «ядерной» коллекции WoS Core Collection был выделен основной массив публикаций, связанных с большими данными на основе на основе следующего поискового образа:

“big data\*” OR bigdata OR “large dataset\*” OR “massive data\*” OR “data science” OR “data\* mining” OR “datamining” OR “text mining” OR “Hadoop\*” OR “MapReduce” OR “Map Reduce” OR “unstructured data\*” OR “semistructured data\*” OR “semi-structured data\*” OR “data analytic\*” OR “descriptive analytic\*” OR “diagnostic analytic\*” OR “predictive analytic\*” OR “prescriptive analytic\*”

Подробности формирования приведенного выше поискового образа описаны в статье [9]. Поиск велся за период с 2016 по 2020 год по таким полям, как «Название», «Аннотация», «Автор» и «Ключевые слова» (применено поле «Тема» в инструментарии InCites платформы WoS), что в итоге дало в качестве результата немногим более 100 000 публикаций.

На следующем этапе из основного массива статей осуществлялся отбор публикаций, посвященных телекоммуникационной инфраструктуре, вычислительной инфраструктуре и инфраструктуре цифровых платформ. Для этого были сформированы специализированные поисковые образы:

- для телекоммуникационной инфраструктуры: “digital infrastructure” OR “information infrastructure” OR “ICT-infrastructure” OR “broadband” OR “broadband fixed” OR “broadband mobile”;
- для вычислительной инфраструктуры: “high-performance computing” OR “supercomputing” OR “cloud computing” OR “cloud service\*” OR IaaS OR SaaS OR PaaS OR DaaS;
- для цифровых платформ: “digital platform\*” OR “transaction platform\*” OR “online platform\*” OR “integrated platform\*”.

Ключевые слова для уточняющих поисковых образов подбирались с целью максимально отразить исследуемую предметную (под)область и одновременно минимизировать информационный «шум» в форме публикаций, слабо связанных с исследуемой (под)областью.

Общий и специализированные поисковые запросы были объединены логическим оператором «AND», в результате поиска по полю «Тема» было найдено 1201 публикация за последние 5 лет. Распределение источников по годам показывает достаточно стабильный исследовательский интерес к цифровой инфраструктуре для работы с большими данными: за 2016 год найдено 167 публикаций, за 2017 г. – 226, за 2018 г. – 266, за 2019 г. – 283, за 2020 г. – 235.

На заключительном этапе был проведен углубленный анализ массива найденных публикаций: на первой стадии изучались аннотации, а затем в случае релевантности публикации – полный текст.

Телекоммуникационной инфраструктуре и широкополосному доступу для работы с большими данными посвящено 13 релевантных статей, в частности публикации, связанные с использованием широкополосного доступа в интернет (далее – ШПД) при работе с большими данными с помощью технологий интернета вещей [17, 18] и искусственного интеллекта [19, 20].

Вычислительной инфраструктуре для работы с большими данными, прежде всего, суперкомпьютерингу посвящено 20 публикаций, в том числе использованию высокопроизводительных вычислений для фундаментальных исследований или прогнозной аналитики [21, 22, 23, 24]. Особое место занимают публикации, посвященные сервисам облачных вычислений – общим числом 727, что демонстрирует широкую распространенность облачных сервисов при работе с большими данными.

Наконец, цифровым платформам для работы с большими данными посвящены 23 научные публикации, затрагивающие вопросы использования социальных сетей [25], анализа политических предпочтений пользователей [25, 26], изучения клиентского поведения [27], сферы образования [28, 29, 30, 31], здравоохранения [32, 33, 34, 35], интернета вещей [36, 37], культуры [38].

Анализ показал, что в отобранных публикациях при мониторинге телекоммуникационной инфраструктуры затрагиваются такие аспекты, как значимость фиксированного и мобильного ШПД для граждан и организаций, перспективы внедрения мобильной связи пятого поколения, а также экономическая доступность. Для развитых стран одной из серьезных проблем является ограниченная пропускная способность существующих каналов связи как внутри страны, так и межстрановых соединений.

При мониторинге вычислительной инфраструктуры для работы с большими данными исследователи делают акцент на потребности в хранении и обработке постоянно возрастающих объемов данных, а также на необходимости иметь в стране или отдельной сфере деятельности универсальные или специализированные вычислительные комплексы, такие как суперкомпьютеры или кластеры для параллельных вычислений.

В отношении оценки платформенной инфраструктуры консенсус по системе показателей для мониторинга пока не сформировался, однако большинство исследователей отмечает важность наличия цифровых платформ для трансформации практически во всех сферах и особенно при работе с большими данными. Важными трендами последних лет стало как появление специализированных цифровых платформ для конкретной сферы деятельности (здравоохранение, сельское хозяйство, транспорт, финансы и т.д.), так и объединение цифровых платформ в экосистемы и даже конкуренция цифровых экосистем между собой.

### **3 Концептуальная схема и показатели мониторинга**

#### **3.1 Концептуальная схема предметной области**

Результаты проведенного анализа приводят к концептуальной схеме мониторинга цифровой инфраструктуры для работы с большими данными, состоящей из трех компонентов: телекоммуникационной инфраструктуры для передачи больших массивов данных, вычислительной инфраструктуры для хранения и обработки больших данных, платформенной инфраструктуры для оперирования большими данными (см. рисунок 1). Важно отметить, что выбор соответствующих компонентов подтверждается одним из первых исследований по анализу текущего состояния развития цифровой экономики в России [15], где соответствующие компоненты также составляют основу для мониторинга цифровой инфраструктуры.

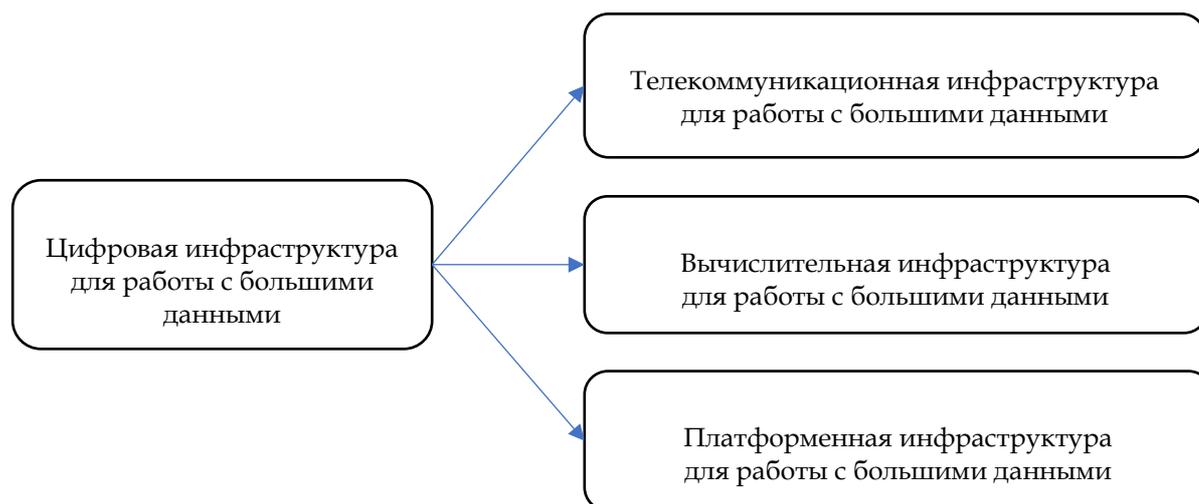


Рис. 1. Концептуальная схема мониторинга и оценки состояния цифровой инфраструктуры для работы с большими данными

Первый компонент предметной области мониторинга цифровой инфраструктуры для работы с большими данными связан с оценкой наличия телекоммуникационной инфраструктуры (страны, региона или сферы деятельности) и ее доступности для основных заинтересованных сторон. Важнейшими характеристиками телекоммуникационной инфраструктуры является скорость и полоса пропускания в сетях передачи данных, а также ее доступность, в том числе – экономическая. В рамках настоящего исследования критически важной является, прежде всего, такая характеристика телекоммуникационной инфраструктуры, как широкополосный доступ, обеспечивающий возможность использования инструментов распределенной работы с большими данными.

Не менее важным фактором, обеспечивающим возможность работы с большими данными, является наличие вычислительной инфраструктуры, предназначенной для хранения и обработки больших массивов данных и предоставляющей как вычислительные мощности центров обработки данных, так и разнообразный спектр сервисов облачных вычислений для конечных пользователей. Наличие в стране, регионе или конкретной сфере деятельности вычислительных ресурсов позволяет не только использовать современные технологии и инструменты работы с большими данными, но и поддерживать отечественных производителей и поставщиков цифровых решений и сервисов, формируя спрос в цифровом секторе экономики. При этом доступность (в том числе экономическая) вычислительных ресурсов для различных категорий пользователей является важной предпосылкой для использования технологий работы с большими данными.

Как уже отмечалось, последние десятилетия характеризуются появлением и доминированием онлайн-цифровых платформ, ставших одним из главных драйверов цифровой трансформации во многих сферах деятельности. При этом сформировалось несколько направлений развития цифровых платформ, связанных как со способами ведения бизнеса, так и со стеком используемых в платформе цифровых технологий [4]. Для целей настоящего исследования существенной является оценка наличия и доступности в стране или конкретной сфере деятельности одной или нескольких цифровых платформ, позволяющих принципиально изменить способ ведения бизнеса и использовать большие данные.

Мониторинг цифровой инфраструктуры для работы с большими данными представляет интерес не только для международных сопоставлений на уровне стран, но и для оценки ситуации в отдельных сферах деятельности. В рамках настоящего исследования рассматриваются приоритетные сферы деятельности, связанные с национальными целями развития Российской Федерации до 2030 года [39]:

- 1) государственное управление;
- 2) здравоохранение;
- 3) наука;

- 4) образование;
- 5) промышленность;
- 6) развитие городской среды;
- 7) сельское хозяйство;
- 8) строительство;
- 9) транспорт и логистика;
- 10) финансовые услуги;
- 11) экология и природопользование;
- 12) энергетическая инфраструктура.

### 3.2 Телекоммуникационная инфраструктура для работы с большими данными

Первый компонент системы мониторинга цифровой инфраструктуры для работы с большими данными должен обеспечить возможность оценки наличия телекоммуникационной инфраструктуры и ее доступности, прежде всего, ШПД. Для целей настоящего исследования будут в первую очередь использоваться согласованные на международном уровне показатели, характеризующие ШПД для организаций, а также экономическую доступность подобных соединений [40].

Не менее важно оценить наличие в стране надежной телекоммуникационной инфраструктуры для использования гражданами, которые, с одной стороны, выступают в качестве потенциальных клиентов для бизнеса, использующего технологии работы с большими данными, а с другой стороны – производят большое количество пользовательских данных, необходимых для разработки и предоставления персонализированных сервисов. Для эффективного решения задач подобного рода необходимы развитая телекоммуникационная инфраструктура сбора и передачи данных, равно как и повсеместный доступ граждан к таким сервисам с другой.

Первую группу характеристик образуют показатели наличия телекоммуникационной инфраструктуры в стране.

- (ТИБД-01-01) Доля организаций, имеющих ШПД к интернету, в общем числе обследованных организаций

Данный показатель характеризует уровень зрелости телекоммуникационной инфраструктуры широкополосного (как фиксированного, так и мобильного) доступа к интернету, что необходимо для массового использования технологий работы с большими данными.

- (ТИБД-01-02) Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, в общем числе обследованных организаций

Высокие скорости доступа к интернету дают преимущества организациям перед конкурентами с меньшей скоростью, так как они имеют возможность быстрее получать, обрабатывать и извлекать «добавленную ценность» из больших массивов данных.

- (ТИБД-01-03) Доля населения, охваченного услугами подвижной связи не ниже 3G

При расчете показателя учитывается население, охваченное услугами сотовой связи 3-го и 4-го поколений. Основные области применения технологий для работы с большими данными – пользовательские данные, которые в основном генерируются мобильными устройствами пользователей

- (ТИБД-01-04) Доля населения, использующего мобильные устройства для доступа к интернету

Распространенность мобильных устройств (таких как смартфоны и планшеты) служат дополнительной характеристикой к показателю ТИБД-01-03, позволяющей оценить возможности для генерации, сбора и обработки пользовательских данных, которых лишены обычные мобильные телефоны.

- (ТИБД-01-05) Международная пропускная способность каналов связи на 1 пользователя

Данный показатель важен в силу того, что совокупная пропускная способность всех международных каналов связи является ограниченной величиной и его усредненная величина на каждого пользователя показывает скорость связи пользователя с внешним миром.

- (ТИБД-01-06) Стоимость корзины услуг фиксированного ШПД к интернету

Доступность ШПД заключается не только в наличии каналов для передачи данных, но и в экономической возможности населения оплачивать доступ в интернет. Поэтому МСЭ уже

достаточно длительное время мониторит стоимость корзины услуг фиксированного ШПД к интернету как плату за ежемесячное подключение к интернету (с ограничением объема трафика в 5 Гб), рассчитанную как доля от среднемесячного валового национального дохода (ВНД) страны на душу населения.

- (ТИБД-01-07) *Стоимость корзины услуг мобильного ШПД к интернету*

Аналогичный предыдущему показатель экономической доступности мобильного ШПД отражает стоимость ежемесячного использования мобильной связи с активным использованием мобильного (безлимитного) интернета, рассчитанную как доля от среднемесячного ВНД страны на душу населения.

### 3.3 Вычислительная инфраструктура для работы с большими данными

Как отмечалось ранее, работа с большими данными требует наличия высокопроизводительных вычислительных систем, способных обеспечить хранение и обработку больших массивов разнообразных данных, генерируемых (или собираемых) с большой скоростью. До недавнего времени для этого хватало традиционных ИКТ-решений, однако быстрое накопление огромных массивов цифровых данных в последнее десятилетие привело к необходимости создания и применения высокопроизводительных систем с принципиально другой архитектурой и производительностью. Наличие в стране (или отдельной сфере деятельности) подобных вычислительных ресурсов и сервисов на их основе открывает возможности для высокопроизводительной аналитики больших данных и моделирования сложных процессов.

Владение подобными высокопроизводительными вычислительными системами предполагает существенные расходы на содержание физической инфраструктуры центров обработки данных с большим количеством компьютерного оборудования, расходы на электроэнергию, обеспечение информационной и физической безопасности, бесперебойности и многое другое. Это далеко не всегда экономически выгодно даже для очень крупных компаний, поэтому выходом может стать совместное использование вычислительных ресурсов, предоставляемых по модели облачных вычислений. Поэтому в систему мониторинга вычислительной инфраструктуры для работы с большими данными включены не только показатели наличия соответствующих вычислительных мощностей в стране, но и характеристики сервисов облачных вычислений.

- (ВИБД-02-01) *Суммарная доля мощностей суперкомпьютеров страны в мировом рейтинге TOP500*

Для оценки уровня вычислительных ресурсов выбрана общая вычислительная мощность всех суперкомпьютеров страны из мирового рейтинга TOP500, отнесенная к размеру экономики, что косвенно характеризует обеспеченность национальной экономики вычислительными ресурсами, в том числе для работы с большими данными.

- (ВИБД-02-02) *Мощность коммерческих центров обработки данных*

В бизнес-среде достаточно распространенным показателем развитости рынка услуг центров обработки данных служит количество стойко-мест в отдельно взятом центре. При всей неоднозначности данного показателя он, тем не менее, может служить качественной характеристикой наличия в стране или отдельно взятой сфере деятельности вычислительных ресурсов для работы с большими данными.

- (ВИБД-02-03) *Доля объема рынка облачных вычислений от мирового*

Данный показатель в большей степени характеризует возможности национальной системы облачных вычислений и экономическую доступность облачных сервисов.

- (ВИБД-02-04) *Доля организаций, использующих сервисы облачных вычислений в общем числе обследованных организаций*

Востребованность и доступность сервисов облачных вычислений, особенно в разрезе по видам облачных сервисов (IaaS, PaaS, SaaS, DaaS и т.д.), открывают возможность для мониторинга и оценки облачной инфраструктуры как в масштабах страны, так и отдельной сферы деятельности.

Данные для показателя ВИБД-02-01 публикуются дважды в год на веб-сайте рейтинга TOP 500 [41]; сведения для расчета показателей ВИБД-02-02 и ВИБД-02-03 регулярно публикуются авторитетными консалтинговыми компаниями (см. например [3], [42] и могут быть рассчитаны в ходе кабинетного исследования; данные для расчета показателя ВИБД-02-04 ежегодно публикуются

по итогам официального статистического наблюдения как в Российской Федерации [43], так и в Европейском Союзе [44], что позволяет проводить международные сопоставления.

### 3.4 Платформенная инфраструктура для работы с большими данными

Современные цифровые платформы построены на работе с (большими) данными как потребителей и поставщиков, так и технологических операторов. Объединение разнообразных цифровых платформ в экосистемы еще в большей степени способствует взаимодействию между различными группами пользователей, предоставляя цифровые сервисы для проведения транзакций, сбора, обработки и совместного использования данных, связанных с их общими интересами или деятельностью.

Для оценки платформенной инфраструктуры в стране будут использованы следующие группы показателей: наличие универсальных цифровых платформ (например, платформ электронной коммерции, поисковых систем, контент-платформ, социальных сетей); наличие специализированных цифровых платформ для отдельных сфер деятельности; наличие технологических платформ, позволяющих использовать тот или иной набор цифровых технологий для работы с данными (например, машинного обучения или интернета вещей).

- (ПИБД-03-01) *Наличие доступных универсальных цифровых платформ*

Данный показатель характеризует наличие доступа пользователей (в том числе экономически приемлемого) к сервисам универсальных цифровых платформ как глобальных (при условии их локализации), так и национальных. При этом успешное функционирование подобных платформ, с одной стороны, порождает огромные массивы пользовательских данных, а с другой – требует новых методов их обработки, выступая драйвером развития технологий работы с большими данными.

- (ПИБД-03-02) *Наличие доступных специализированных цифровых платформ для отдельных сфер деятельности*

Включение данного показателя в систему мониторинга и оценки платформенной инфраструктуры открывает возможность для сравнительного анализа ситуации в разрезе по различным сферам деятельности как внутри страны, так и для международных сопоставлений.

- (ПИБД-03-03) *Наличие технологических цифровых платформ*

Показатель нацелен на мониторинг и оценку уровня технологического развития цифровых платформ в стране и для проведения международных сопоставлений. Его также можно рассматривать в разрезе отдельных сфер деятельности и оценивать их технологическую оснащенность для работы с данными.

- (ПИБД-03-04) *Наличие экосистемы цифровых платформ*

Использование данного показателя предоставит возможности мониторинга дальнейшей эволюции платформенной инфраструктуры, прежде всего, с точки зрения проникновения единых экосистем в различные сферы деятельности, равно как и объединения в единой экосистеме стека разнообразных цифровых технологий работы с данными. При этом важно отслеживать не только наличие экосистем в стране, но и их доступность (включая экономическую) и уровень конкуренции.

Дальнейшее развитие предложенной системы показателей для платформенной инфраструктуры может идти в направлении учета аспектов, связанных с повышением уровня зрелости как универсальных/специализированных платформ, так и технологических цифровых платформ. Аналогичные показатели можно ввести и по отношению к экосистемам цифровых платформ.

## 4 Результаты пилотной реализации

В настоящее время многие показатели мониторинга цифровой инфраструктуры для работы с большими данными можно рассчитать методом кабинетных исследований на основе открытых данных официальных статистических наблюдений. Ниже приведены результаты, основанные на последних доступных данных по итогам 2019 года. В ряде случаев показатели рассчитаны для Российской Федерации в целом, а в некоторых случаях – для отдельно взятых сфер деятельности. Всюду, где возможно, даются международные сопоставления Российской Федерации со странами Европейского союза.

#### 4.1 Телекоммуникационная инфраструктура для работы с большими данными

В 2019 году по показателю «ТИБД-01-01 Доля организаций, подключенных к интернету, в общем числе обследованных организаций» Россия достигла значения 86,6%, что соответствует уровню таких европейских стран, как Венгрия – 86,88% и Латвия 87,91%, меньше значение только у Греции: 81,55% (см. ниже рисунок 2).

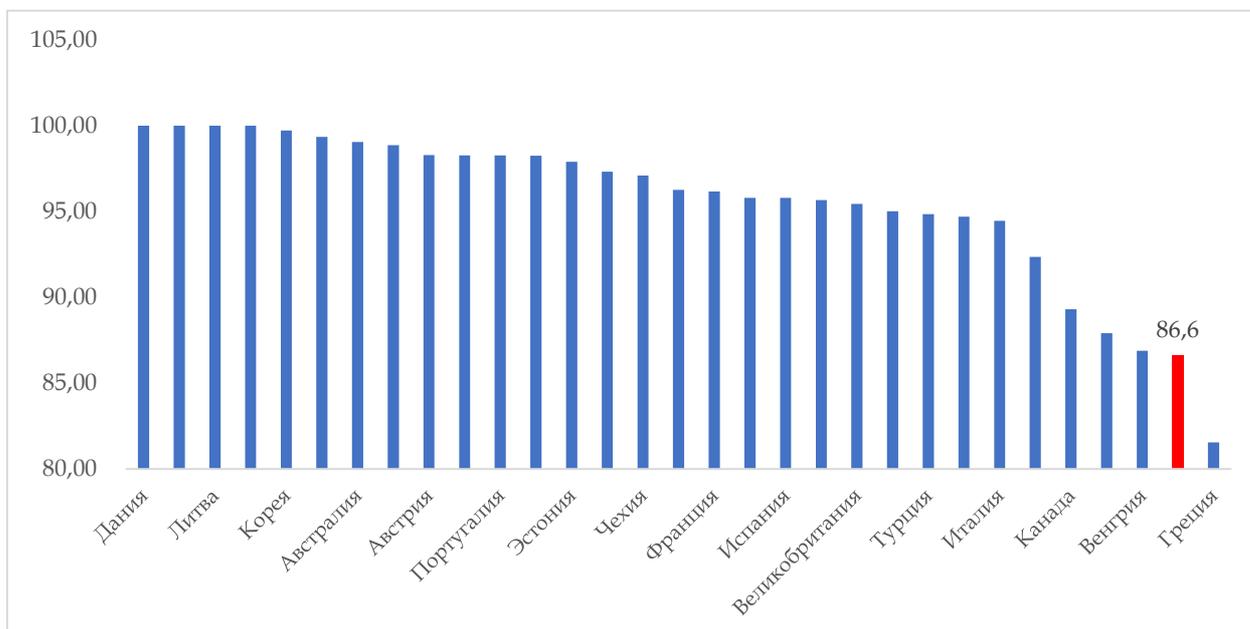


Рис. 2. Доля организаций, подключенных к интернету, %, 2019, ЕС, Россия

Источники: [43] [44]

Следует отметить, что большинство европейских стран перешагнуло 80%-й рубеж еще в 2010 году, а в 2019 среднее значение для стран ЕС составляет 97% и остается неизменным в течение последних пяти лет. В России идет замедляющийся рост количества организаций, имеющих доступ к интернету: например, с 2018 по 2019 год прирост составил всего 0,1%.

Наряду со страновыми показателями интернет-коннективности российских организаций, представляет интерес оценка по этому показателю для различных сфер деятельности (см. рисунок 3).

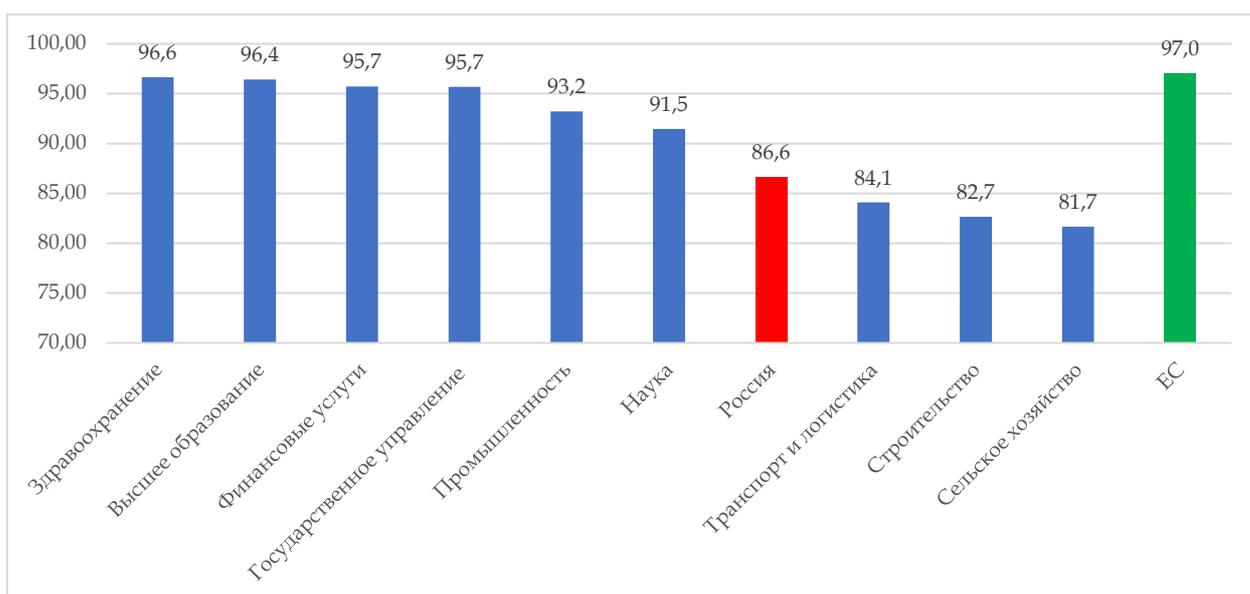


Рис. 3. Доля организаций, подключенных к интернету, %, 2019, в разрезе сфер деятельности,

Источники: [43] [44]

Отраслями-лидерами в Российской Федерации в 2019 году по показателю ТИБД-01-01 являются здравоохранение, высшее образование, финансы, государственное управление, промышленность и наука, для которых значение данного показателя выше среднего значения по России. Для транспорта, строительства и сельского хозяйства значения показателя – ниже среднего. При этом лидирующие сферы деятельности не сильно уступают среднему значению для Европейского союза. Для таких сфер, как развитие городской среды и энергетическая инфраструктура, данные в открытых источниках не доступны.

По показателю (ТИБД-01-02) «Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек» (см. рисунок 4) Россия существенно уступает странам ЕС обеспечивая в 2019 году ШПД на скорости не менее 30 Мбит/сек примерно 34% организаций, тогда как страны-лидеры достигли значений более 60% (Дания, Швеция, Португалия, Люксембург, Финляндия).

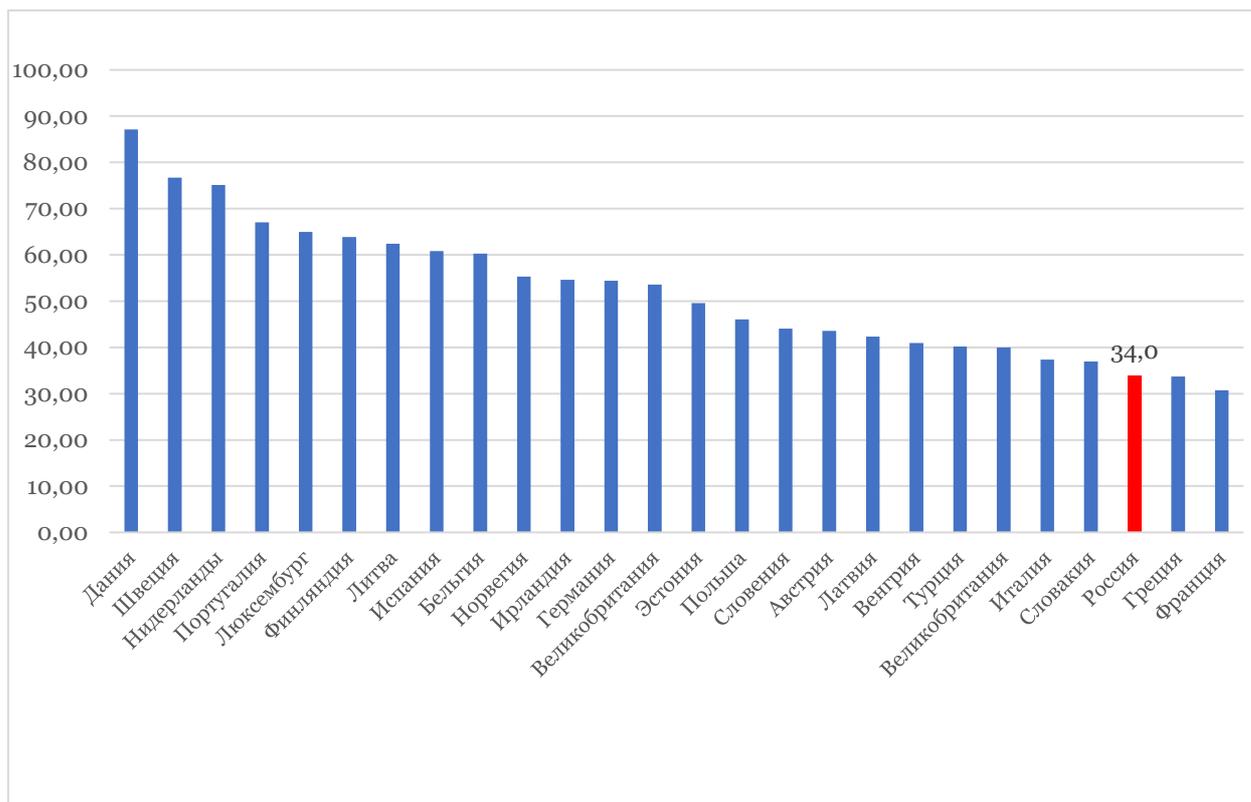


Рис. 4. Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, %, 2019, ЕС, Россия

Источники: [43] [44]

Динамика изменения показателя (ТИБД-01-02) последние 5 лет демонстрирует устойчивый рост для всех стран без исключения. Страны-лидеры ЕС при этом в среднем прибавляют в год по 5 процентных пунктов, что более чем в два раза превышает рост в Российской Федерации. Такими темпами европейские страны-лидеры решат задачу обеспечения организаций телекоммуникационной инфраструктурой в ближайшие 5–7 лет, тогда как России при сохранении имеющихся темпов для достижения сопоставимого уровня понадобится 30–40 лет.

В то же время ситуация с ШПД для организаций из отдельных сфер деятельности сильно различается: высшее образование, наука, промышленность и здравоохранение существенно превышают средние российские значения, тогда как система государственного управления и сельское хозяйство значительно уступают (см. рисунок 5). Особо следует отметить сферу высшего образования, где ШПД заметно превышает даже общеевропейский уровень.

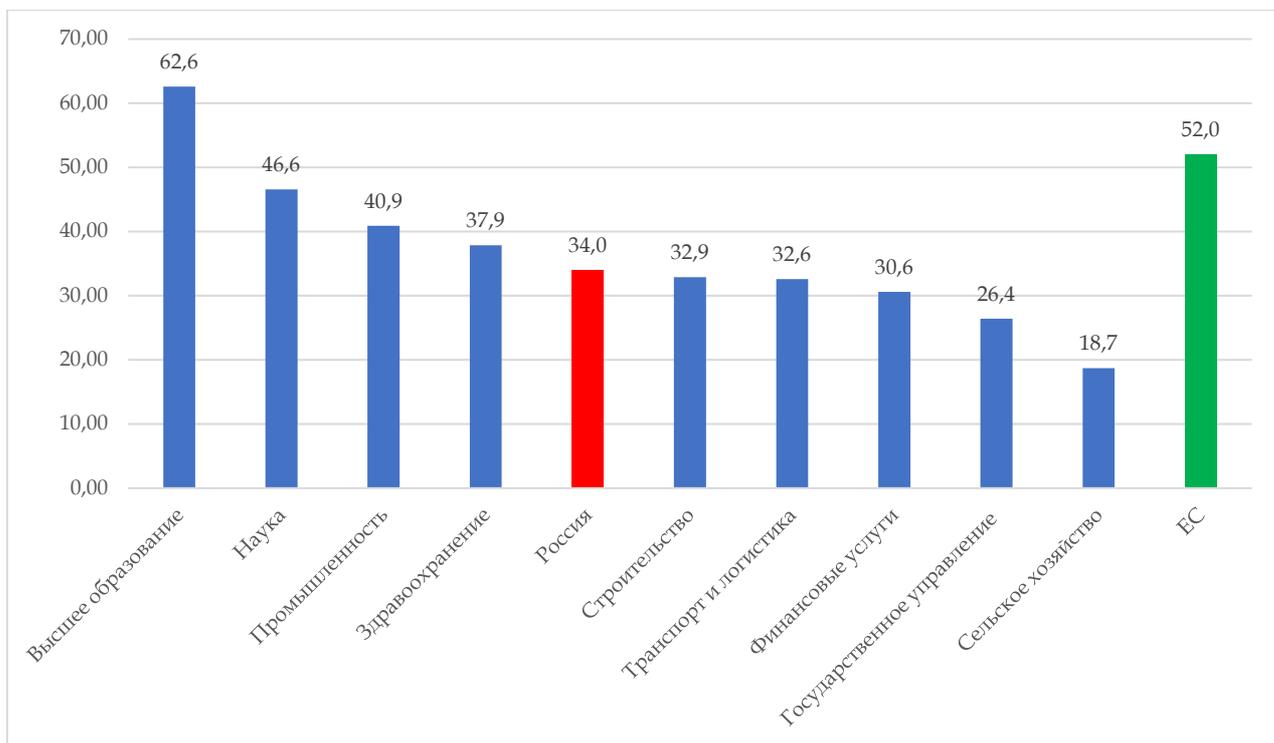


Рис. 5. Доля организаций, имеющих скорость передачи данных через интернет не менее 30 Мбит/сек, %, 2019, в разрезе сфер деятельности, Источники: [43] [44]

Охват населения услугами подвижной связи третьего поколения и выше характеризуется показателем ТИБД-01-03, по которому в 2019 году Россия достигла значения 87,7% (см. рисунок 6), тогда как для стран или страновых объединений, сопоставимых с Россией по размерам (ЕС, США, Китай, Бразилия, Канада), эта задача практически полностью решена (разве что кроме Бразилии).

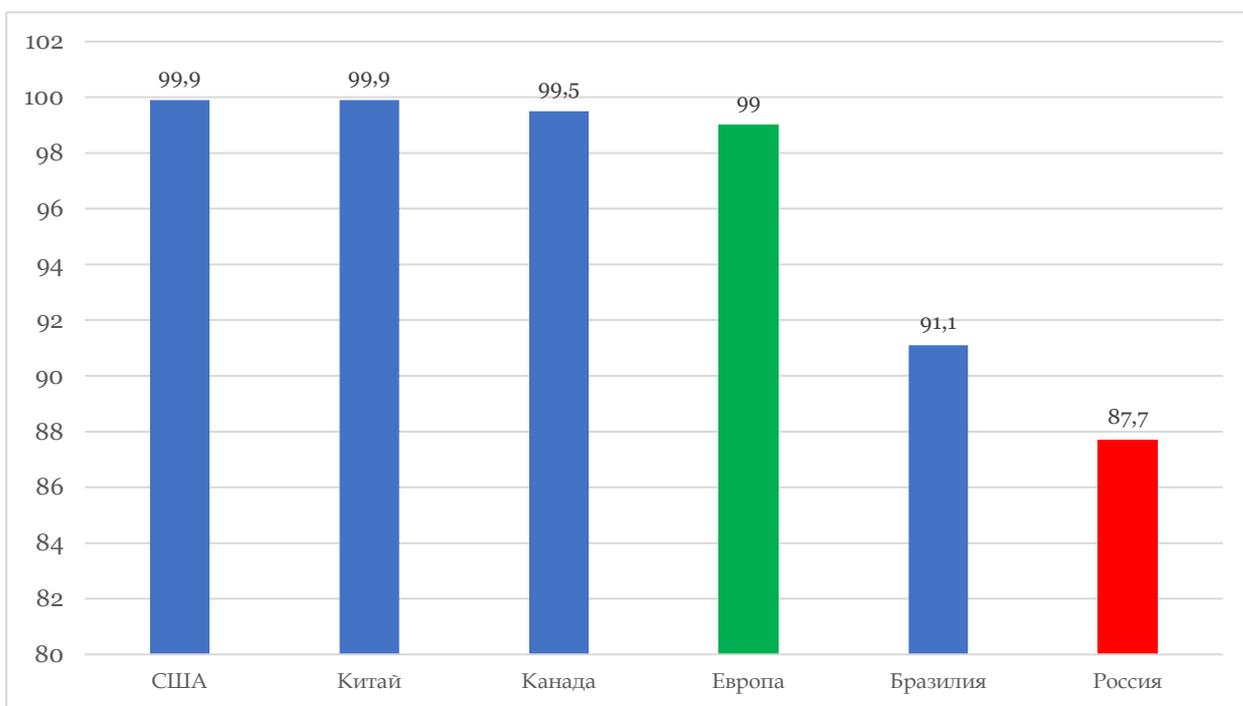


Рис. 6. Охват населения услугами подвижной связи не ниже 3G, %, 2019, Источники: [45]

В 2018 году 73,5% населения России использовало телефон или смартфон для доступа к интернету, что соответствует среднему уровню ЕС – 72% (см. рисунок 7), при этом к 2020 году наблюдалось снижение этих показателей [46].

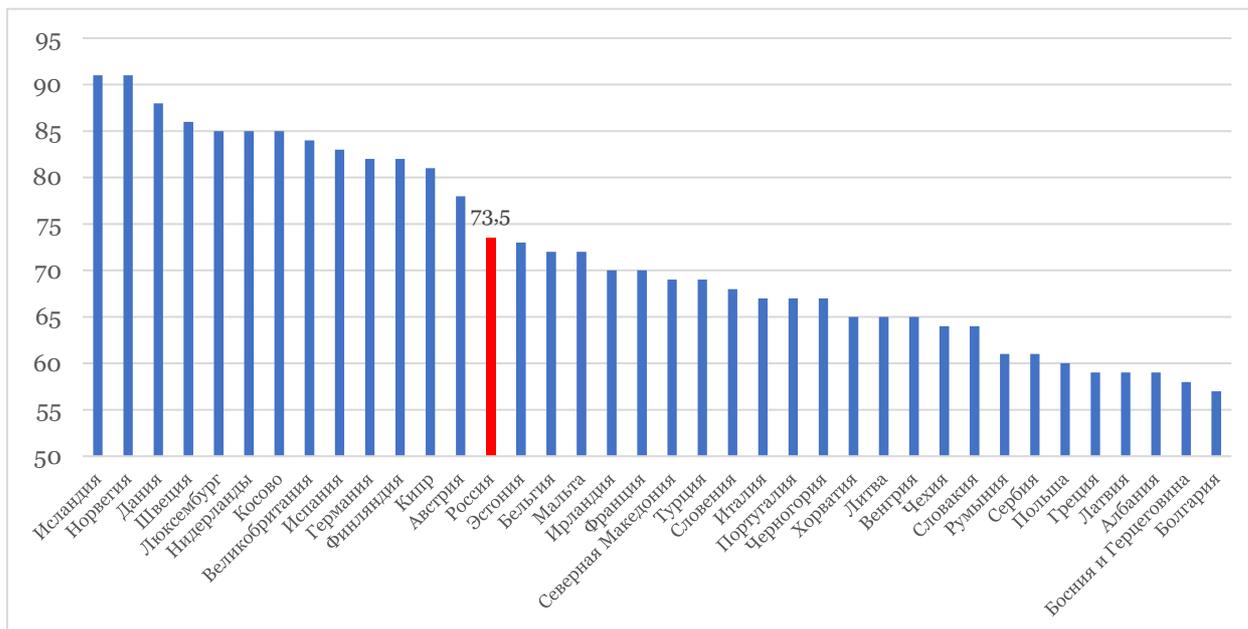


Рис. 7. Использование населением телефона или смартфона для доступа к интернету, %, 2018, Россия, ЕС  
Источники: [44], [46]

По показателю (ТИБД-01-05) «Международная пропускная способность каналов связи на 1 пользователя» в 2019 году РФ со значением в 62 842,6 Кбит/с находится в конце рейтинга, опережая такие страны, как Тунис, Казахстан, Монголия, Египет и Куба. Лидерами являются Гонконг (11 614 411 Кбит/с) и Кипр (1 178 121 Кбит/с). Ввиду более чем 180-кратного превышения значения лидера (Гонконга) относительно России, на графике он не приводится (см. рисунок 8).

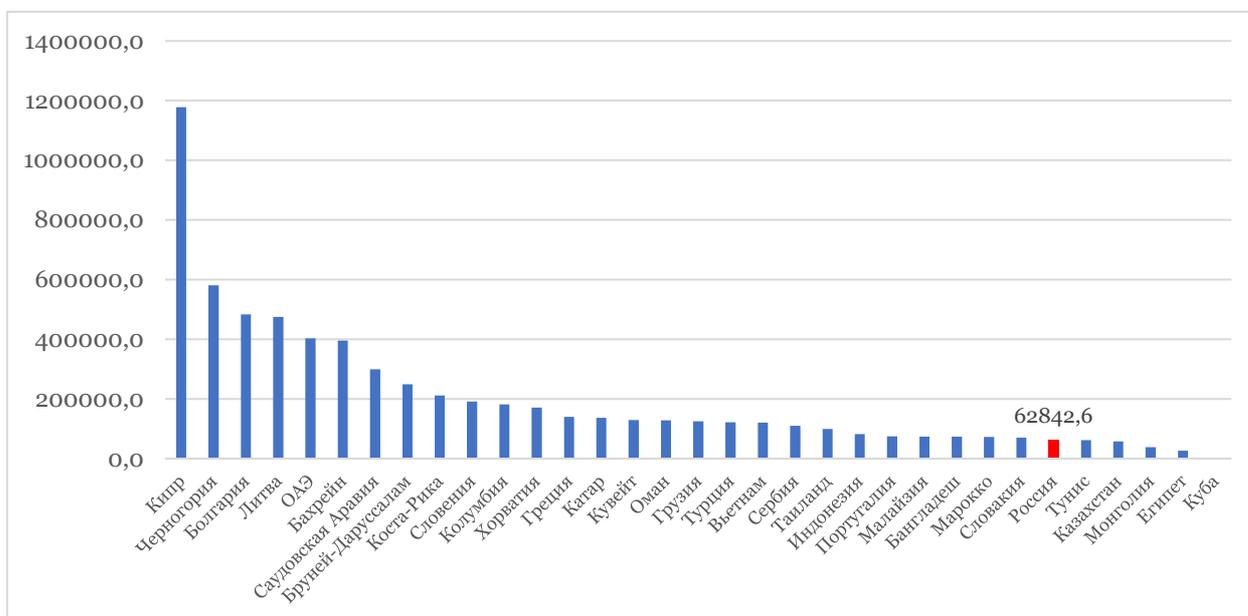


Рис. 8. Международная пропускная способность каналов связи на 1 пользователя  
Источники: [45]

Динамика изменения международной пропускной способности стран показывает, что с 2015 по 2019 год страны-лидеры показывали значительный рост, кроме Болгарии в 2018 году. Гонконг в среднем за год прибавлял по 1 357 801,8 Кбит/с, Кипр – более чем 220 000 Кбит/с, другие европейские страны-лидеры – от 66 000 Кбит/с (Болгария и Литва) до 97 000 Кбит/с (Черногория).

Россия существенно отстает по темпам роста, с 2015 по 2017 год демонстрируя рост, который затем сменился спадом в 2018 и 2019 годах на 4 000 и 2 000 Кбит/с соответственно.

По показателям экономической доступности в 2020 году Россия вошла в десятку стран-лидеров по стоимости услуг фиксированного ШПД и заняла 8-е место со значением в 0,64% (ВНД на душу населения) из 178 стран (см. рисунок 9). Лидерами рейтинга являются Лихтенштейн, Кувейт, Китай и Объединенные Арабские Эмираты; аутсайдерами рейтинга являются Малави, Нигер и Мадагаскар, где стоимость услуг фиксированного ШПД превышает 100% ВНД на душу населения.

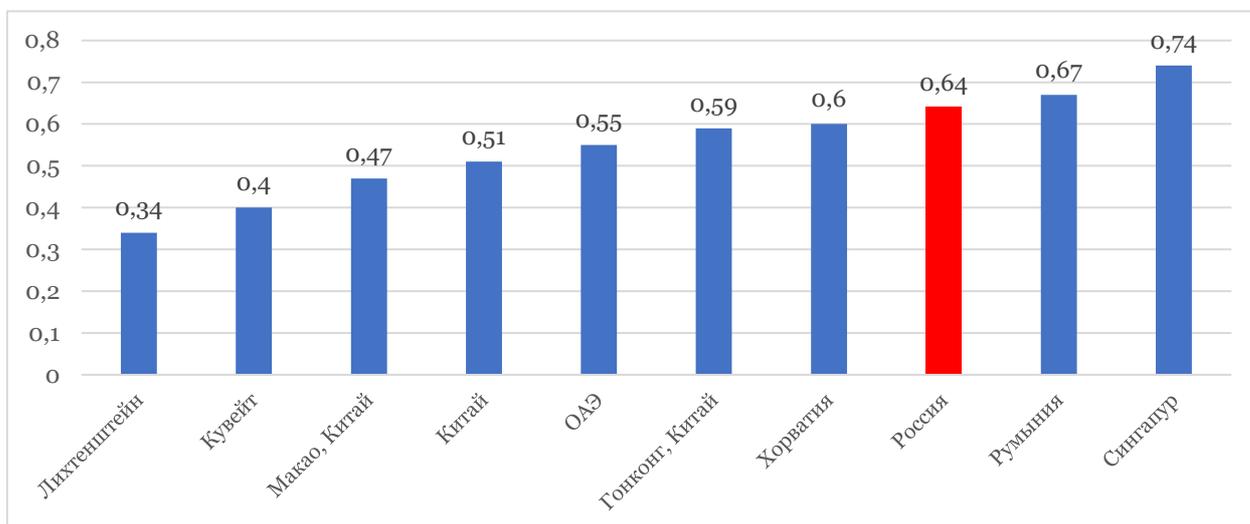


Рис. 9. Стоимость фиксированного ШПД в 2020 г., % ВНД на душу населения  
Источники: [47]

По стоимости услуг мобильного ШПД в 2020 году Россия входит в двадцатку стран-лидеров и занимает 19-е место со значением в 0,67% (ВНД на душу населения) из 178 стран (см. рисунок 10). Лидерами являются Лихтенштейн, Люксембург, Австрия, Китай и Израиль со значениями до 0,3% (ВНД на душу населения); хуже всего дело обстоит в таких африканских странах, как Чад, Центральноафриканская Республика, Бурунди, Малави и Демократическая Республика Конго, где стоимость услуг мобильного ШПД превышает 50% (ВНД на душу населения).

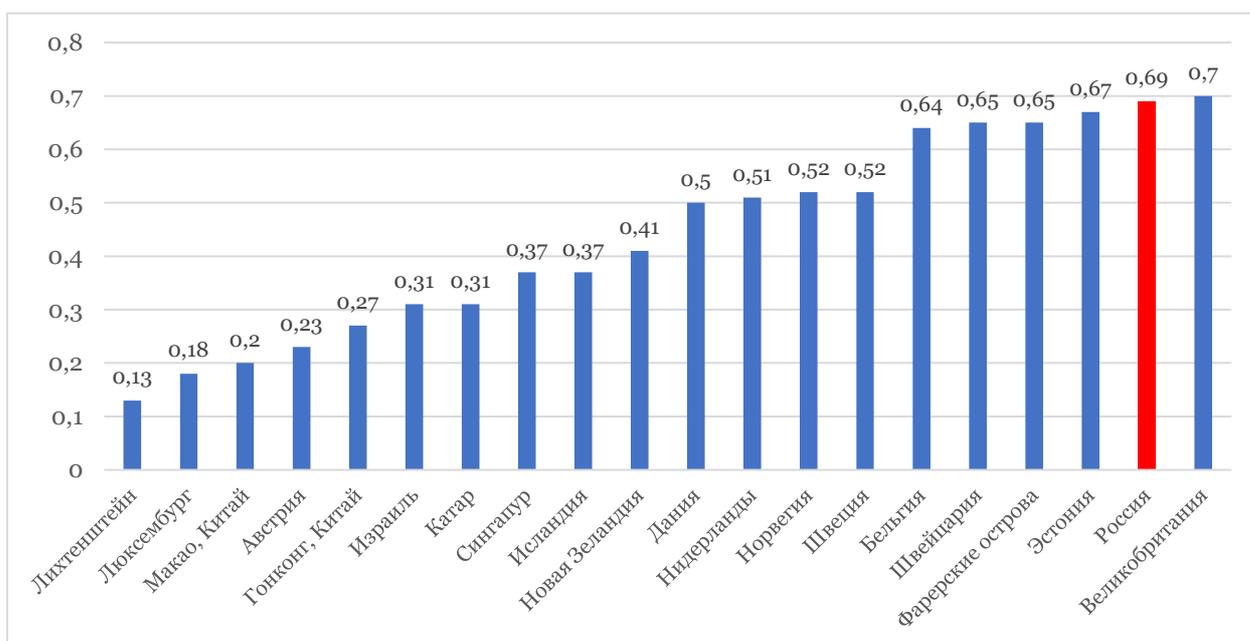


Рис. 10. Стоимость мобильного ШПД в 2020 г., % ВНД на душу населения  
Источники: [47]

## 4.2 Вычислительная инфраструктура для работы с большими данными

Как уже отмечалось в разделе 3.3, одним из показателей наличия в стране вычислительных ресурсов для работы с большими данными служит суммарная доля мощностей суперкомпьютеров в рейтинге TOP 500 (показатель ВИБД-02-01).

Анализ рейтинга показал, что в 2019 году доля России по этому показателю составляла всего 0,63% что соответствует вычислительной мощности в 10347,3 Терафлопс. В 2020 году по сравнению с 2019 годом значение показателя ВИБД-02-01 еще уменьшилось и в рейтинге TOP 500 осталось всего 2 российских суперкомпьютера с вычислительной мощностью 9147 Терафлопс. Лидером рейтинга по числу компьютеров в ноябре 2020 являлся Китай (212 систем в списке), на втором месте США (113 систем), далее Япония (34 системы). При этом, несмотря на меньшее количество суперкомпьютерных систем, США возглавляют список по совокупной производительности в 668,7 Петафлопс (27,5% от суммарной производительности всех систем рейтинга), у Китая – 566,0 Петафлопс (23,2%). Благодаря лидирующей системе Fugaku Япония с общей производительностью суперкомпьютеров в 593,7 Петафлопс (24,4%) опережает Китай и находится на втором месте в мире (см. рисунок 11).

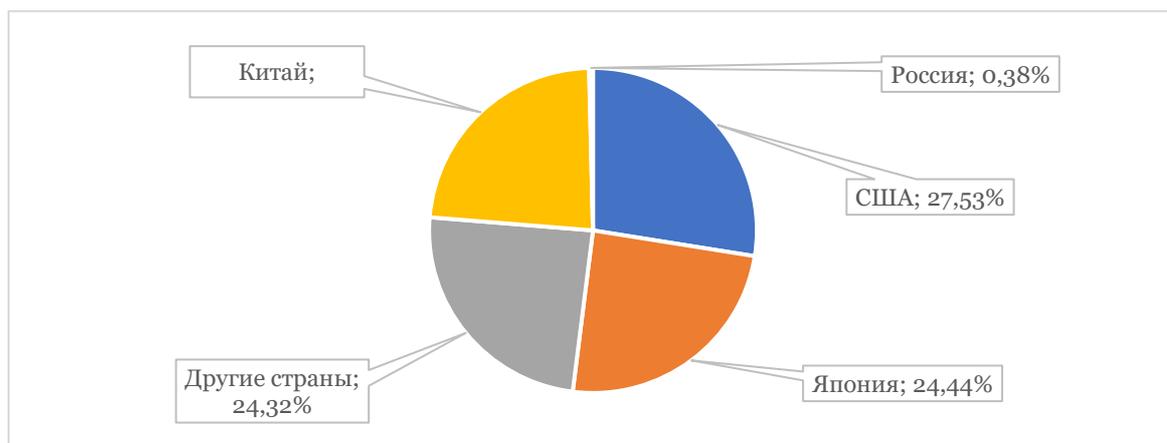


Рис. 11. Доля вычислительной мощности в рейтинге ТОП500 лидеров и России, %, 2020  
Источники: [41]

По использованию сервисов облачных вычислений организациями (показатель ВИБД-02-04) в 2018 году Российская Федерация достигла значения в 26%, что соответствует среднему уровню ЕС (26%), лидерами являются скандинавские страны Финляндия – 65%, Швеция – 57%, Дания – 56%, Норвегия – 51% и Нидерланды – 48% (см. рисунок 12). В 2019 году значение Российской Федерации по данному показателю выросло до 28%, в ЕС данные за этот год не собирались.

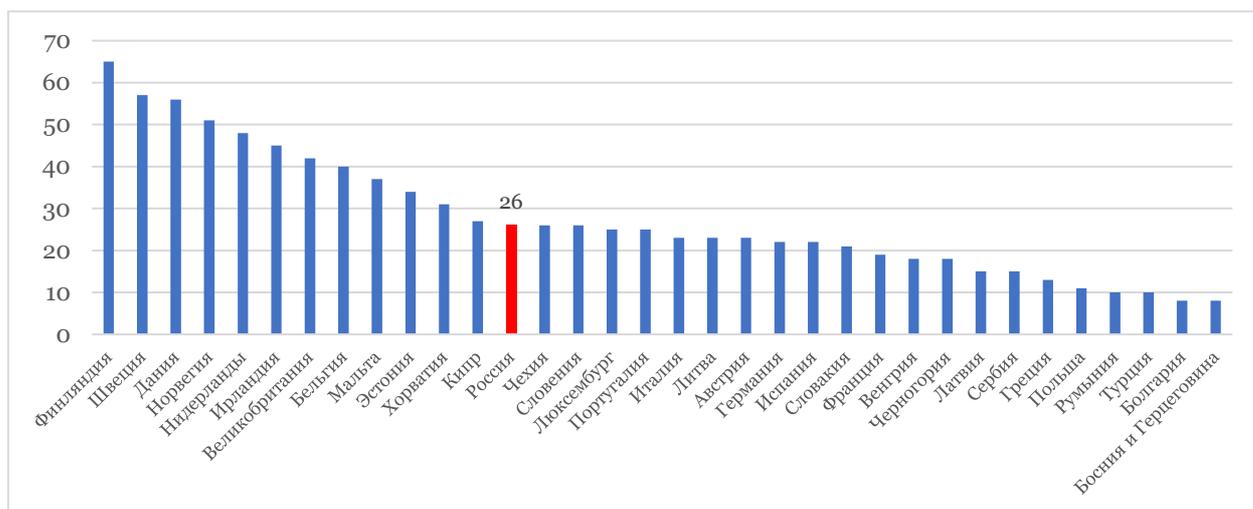


Рис. 12. Доля организаций, использующих сервисы облачных вычислений, %, 2018, ЕС, Россия  
Источники: [43,44]

Отраслевой разрез по доле компаний использующих услуги облачных вычислений в 2019 году показывает, что лидерами среди различных сфер деятельности являются высшее образование (49%), финансовый сектор (38%) и здравоохранение (38%), их значения в том числе превышают и средний уровень для ЕС (30%). Использование услуг облачных вычислений в науке (29%) и промышленности (28%) соответствуют среднему значению по России, тогда как система государственного управления, транспорт, строительство и сельское хозяйство уступают средним общероссийским значениям, но незначительно (см. рисунок 13).

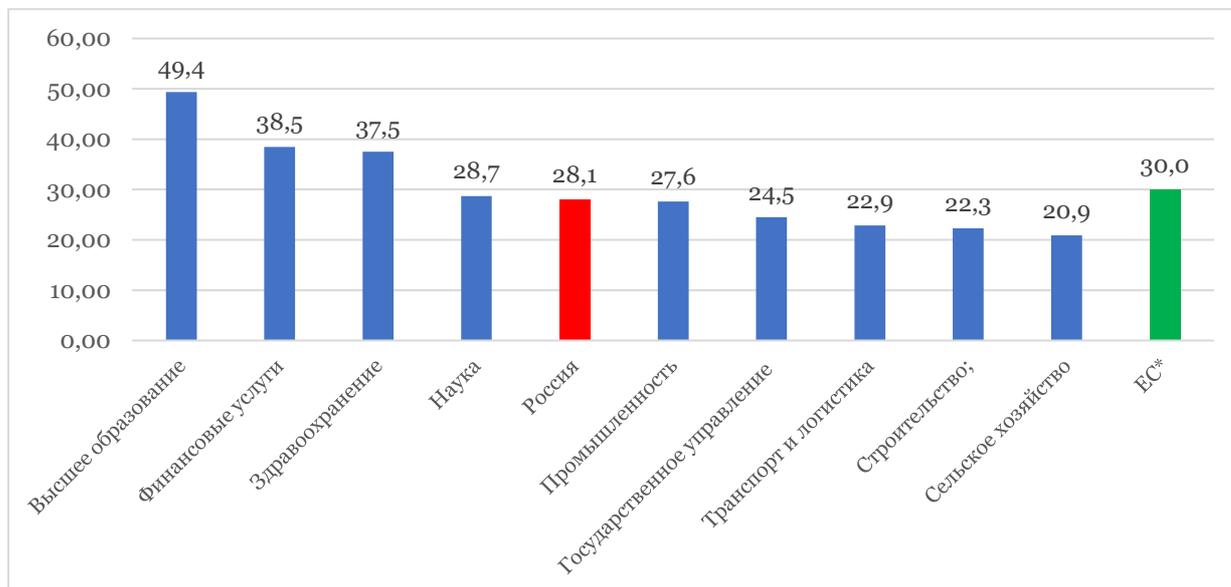


Рис. 13. Доля организаций, использующих сервисы облачных вычислений, %, в разрезе сфер деятельности, Россия (2019), ЕС (2018 – данные в 2019 году не собирались)  
Источники: [43] [44]

#### 4.3 Платформенная инфраструктура для работы с большими данными

Мониторинг платформенной инфраструктуры по предложенной системе показателей в настоящее время в рамках статистического наблюдения только зарождается: первые показатели использования цифровых платформ в 2020 году появились в обновленной форме №3-информ [48], хотя этих сведений для полноценной оценки наличия цифровых платформ различных типов для работы с большими данными явно недостаточно.

Поэтому в рамках системы мониторинга BD4DE предполагается, что данные для расчета показателей (ПИБД-03-01) – (ПИБД-03-04) будут собираться путем экспертных опросов с использованием специального инструментария (экспертной анкеты). Разработанная анкета прошла две стадии пилотирования.

На первой стадии набор показателей по мониторингу платформенной инфраструктуры прошел экспертную апробацию у специалистов по мониторингу развития цифровой экономики и статистики информационного общества. Ими были сформулированы рекомендации по расширению перечня показателей для учета дополнительных факторов, влияющих на применение технологий работы с большими данными. В частности, был предложен дополнительный показатель ПИБД-03-04 «Наличие экосистемы цифровых платформ», а по показателю ПИБД-03-02 «Наличие доступных специализированных цифровых платформ» было предложено собирать данные в разрезе отдельных сфер деятельности.

На второй стадии пилотирования системы мониторинга проводилась апробация анкеты экспертом в данной предметной области (специалистом по применению цифровых технологий в организациях предпринимательского сектора) и экспертом-социологом (специалистом по разработке анкет и проведению опросов организаций), от которых были получены предложения по оптимизации структуры анкеты, снабжении анкеты необходимым понятийным аппаратом и уточнению формулировок отдельных вопросов.

По итогам проведения пилотирования экспертная анкета была доработана и подготовлена для проведения опроса, для повышения эффективности был разработан онлайн-вариант анкеты.

## Заключение

Описанная концептуальная схема и набор показателей мониторинга и оценки цифровой инфраструктуры для работы с большими данными охватывает все уровни: от телекоммуникационной инфраструктуры передачи данных и вычислительной инфраструктуры для их хранения и обработки до платформенной инфраструктуры, предоставляющей возможность использовать результаты оперирования с большими данными в различных сферах деятельности. Некоторые показатели можно рассчитать на основе официальной международной статистики (МСЭ, ОЭСР, Евростат) и результатов федерального статистического наблюдения, регулярно публикуемого Росстатом, однако часть предложенных показателей может быть рассчитана только в результате проведения кабинетных исследований или экспертных опросов.

Важно отметить, что проведение кабинетных исследований для получения необходимых данных может быть затруднено в силу нерегулярного выпуска аналитических материалов, авторами которых часто выступают консалтинговые компании или независимые исследовательские организации, у которых нет публичных обязательств передавать соответствующие исследования в открытый доступ, поэтому часть необходимых данных может либо отсутствовать вовсе за нужный период времени, либо распространяться на коммерческой основе.

В последнее время система статистических наблюдений за цифровой инфраструктурой как со стороны международных организаций разного уровня (МСЭ, ОЭСР, Евростат), так и со стороны России активно модифицируется, в том числе идет активное включение в нее показателей, связанных с работой с большими данными. Дальнейшее развитие системы мониторинга BD4DE, связанное с цифровой инфраструктурой для работы с большими данными, должно учитывать изменения в системе статистических наблюдений международных и российских организаций и оперативного расширения перечня релевантных показателей.

## Благодарности

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

Работа выполнена при частичной поддержке РФФИ, грант 18-29-03086.

## Литература

1. TeleGeography's Global Bandwidth Research Service. URL: <https://www2.telegeography.com/global-bandwidth-research-service> (дата обращения 01.06.2021)
2. Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (дата обращения 01.06.2021)
3. MarketsandMarkets. High Performance Computing (HPC) Market. URL: <https://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html> (дата обращения 01.06.2021)
4. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
5. Still, K., Seppänen, M., Korhonen, H., Valkokari, K., Suominen, A., & Kumpulainen, M. (2017). Business Model Innovation of Startups Developing Multisided Digital Platforms. In Y. Manolopoulos, M. Zaki, D. Diaz, & B. Theodoulidis (Eds.), 2017 IEEE 19th Conference on Business

- Informatics (CBI) (Vol. 2, pp. 70-75). [8012942] IEEE Institute of Electrical and Electronic Engineers. <https://doi.org/10.1109/CBI.2017.86>
6. Rossotto, C. M., Lal Das, P., Gasol Ramos, E., Clemente Miranda, E., Badran, M. F., Martinez Licetti, M., & Miralles Murciego, G. (2018). Digital platforms: A literature review and policy implications for development. *Competition and Regulation in Network Industries*, 19(1–2), 93–109. <https://doi.org/10.1177/1783591718809485>
  7. Yaroslav Eferin, Yuri Hohlov, Carlo Rossotto (2019) «Digital platforms in Russia: competition between national and foreign multi-sided platforms stimulates growth and innovation», *Digital Policy, Regulation and Governance*, Vol. 21, Issue 2, pp. 129-145. URL: <https://doi.org/10.1108/DPRG-11-2018-0065> (дата обращения 01.06.2021); Еферин Я.Ю., Россотто К.М., Хохлов Ю.Е. Цифровые платформы в России: конкуренция между национальными и зарубежными многосторонними платформами стимулирует экономический рост и инновации. *Информационное общество*, 2019, №1–2. С. 16–34. URL: <http://infosoc.iis.ru/article/view/135> (дата обращения 01.06.2021)
  8. Jacobides, MG, Cennamo, C, Gawer, A. Towards a theory of ecosystems. *Strat Mgmt J.* 2018; 39: 2255–2276. <https://doi.org/10.1002/smj.2904>
  9. 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. ??? <https://doi.org/10.1145/3453483.3454078>
  10. ITU. Measuring the Information Society Report. URL: <https://www.itu.int/pub/D-IND-ICTOI> (дата обращения 01.06.2021)
  11. World Telecommunication/ICT Indicators Database. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx> (дата обращения 01.06.2021)
  12. OECD Statistics by theme. URL: <https://stats.oecd.org/> (дата обращения 01.06.2021)
  13. Eurostat Statistics by theme. URL: <https://ec.europa.eu/eurostat/web/main/data/browse-statistics-by-theme> (дата обращения 01.06.2021)
  14. Федеральная служба государственной статистики. Официальная статистика. URL: <https://rosstat.gov.ru/folder/10705> (дата обращения 01.06.2021)
  15. OECD (2019), *An Introduction to Online Platforms and Their Role in the Digital Transformation*, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
  16. Анализ текущего состояния развития цифровой экономики в России. М.: Институт развития информационного общества, 2018. — 166 с. URL: <http://deca.iis.ru> (дата обращения 01.06.2021)
  17. Mossberger Karen, Tolbert Caroline J. Digital Citizenship and Digital Communities: How Technology Matters for Individuals and Communities. <https://doi.org/10.4018/IJEPR.20210701.0a2>
  18. Santos GL, Endo PT, Sadok D; Kelner J. When 5G Meets Deep Learning: A Systematic Review. URL: <https://doi.org/10.3390/a13090208>
  19. Seetharam Karthik, Kagiya Nobuyuki, Sengupta Partho P. Application of mobile health, telemedicine and artificial intelligence to echocardiography. <https://doi.org/10.1530/ERP-18-0081>
  20. Barricelli Barbara Rita; Casiraghi Elena; Fogli Daniela. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. <https://doi.org/10.1109/ACCESS.2019.2953499>
  21. Zacharov Igor, Arslanov Rinat, Gunin Maksim, Stefonishin Daniil, Bykov Andrey, Pavlov Sergey, Panarin Oleg, Maliutin Anton, Rykovanov Sergey, Fedorov Maxim. “Zhores” - Petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. <https://doi.org/10.1515/eng-2019-0059>
  22. Ved Mohit, Rizwanahmed B. Big Data Analytics in Telecommunication using state-of-the-art Big Data Framework in a Distributed Computing Environment: A Case Study. <https://doi.org/10.1109/COMPSAC.2019.00066>
  23. Zacharov I., Panarin O., Ryabinkin E., Izotov K., Teslyuk A. Virtualization for Scientific Workload. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000462194600017> (дата обращения 01.06.2021)
  24. Liao Jianwei, Gerofi Balazs, Lien Guo-Yuan, Miyoshi Takemasa, Nishizawa Seiya, Tomita Hirofumi, Liao Wei-Keng, Choudhary Alok, Ishikawa Yutaka. A flexible I/O arbitration framework for netCDF-based big data processing workflows on high-end supercomputers. URL:

- <https://www.webofscience.com/wos/woscc/full-record/WOS:000405351700005> (дата обращения 01.06.2021)
25. Casero-Ripolles Andreu. Influencers in the Political Conversation on Twitter: Identifying Digital Authority with Big Data. <https://doi.org/10.3390/su13052851>
  26. Sampugnaro Rossana, Montemagno Francesca. In Search of the Americanization: Candidates and Political Campaigns in European General Election. <https://doi.org/10.1080/15377857.2020.1869832>
  27. Al-Abdallah Ghaith, Khair Nadine, Elmarakby Reyad. The Impact of Social Networking Sites on Luxury Vehicles Purchase Decision Process in Gulf Cooperation Council Countries. <https://doi.org/10.1080/08961530.2020.1867023>
  28. Fiofanova Olga Aleksandrovna. New literacy and data-future in education: advanced technology smart big-data. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000572957200014> (дата обращения 01.06.2021)
  29. Yu Jun, Couldry Nick. Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. <https://doi.org/10.1080/1369118X.2020.1764604>
  30. Efimov V, Lapteva A. Digital technology in higher education: situation analysis and prospects assessment (on the example of Krasnoyarsk Krai). <https://doi.org/10.21125/inted.2019.1392>
  31. Larionova Viola, Sheka Andrey, Vasilyev Stanislaus. Students Behavioural Patterns on the National Open Education Platform. <https://doi.org/10.34190/EEL.19.126>
  32. Schiavone Francesco, Leone Daniele, Sorrentino Annarita, Scaletti Alessandro. Re-designing the service experience in the value co-creation process: an exploratory study of a healthcare network. <https://doi.org/10.1108/BPMJ-11-2019-0475>
  33. Pollock Daniel A. Digital Platforms as a Method of Invention for Infection Surveillance. <https://doi.org/10.1089/sur.2019.147>
  34. Aledavood Talayeh, Hoyos Ana Maria Triana, Alakorkko Tuomas, Kaski Kimmo, Saramaki Jari, Isometsa Erkki, Darst Richard K. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. <https://doi.org/10.2196/resprot.6919>
  35. Moller Arlen C., Merchant Gina, Conroy David E., West Robert, Hekler Eric, Kugler Kari C., Michie Susan. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. <https://doi.org/10.1007/s10865-016-9818-7>
  36. Ivaschenko Anton, Korchivoy Stanislav, Spodobaeв Michail. Infrastructural Models of Intermediary Service Providers in Digital Economy. [https://doi.org/10.1007/978-3-030-29513-4\\_44](https://doi.org/10.1007/978-3-030-29513-4_44)
  37. Попов E.V. Econotronics of a smart city. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000476663600010> (дата обращения 01.06.2021)
  38. Kiraly, Peter; Buechler, Marco. Measuring completeness as metadata quality metric in Europeana. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000468499302103> (дата обращения 01.06.2021)
  39. Указ Президента Российской Федерации от 21.07.2020 г. № 424 «О национальных целях развития Российской Федерации до 2030 года». URL: <http://www.kremlin.ru/events/president/news/63728> (дата обращения 01.06.2021)
  40. ITU. Measuring digital development ICT price trends 2020. URL: [https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU\\_ICTPriceTrends\\_2020.pdf](https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceTrends_2020.pdf) (дата обращения 01.06.2021)
  41. TOP 500. URL: <https://www.top500.org/lists/top500/> (дата обращения 15.06.2021)
  42. IDC. Локомотивом развития облачного рынка стали платформенные сервисы. URL: <https://www.idc.com/getdoc.jsp?containerId=prEUR247478021> (дата обращения 15.06.2021)
  43. Федеральная служба государственной статистики. Итоги федерального статистического наблюдения по форме № 3-информ «Сведения об использовании информационных и коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказании услуг в этих сферах». URL: <https://rosstat.gov.ru/folder/14478> (дата обращения 01.10.2021)
  44. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (дата обращения 01.06.2021)
  45. ITU Statistics. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (дата обращения 01.06.2021)

46. Росстат. Итоги выборочного федерального статистического наблюдения по вопросам использования населением информационных технологий и информационно-телекоммуникационных сетей. URL: [https://gks.ru/free\\_doc/new\\_site/business/it/ikt20/index.html](https://gks.ru/free_doc/new_site/business/it/ikt20/index.html) (дата обращения 01.06.2021)
47. ITU. The historical data for the price baskets for countries and regions. URL: [https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU\\_ICTPriceBaskets\\_2008-2020.xlsx](https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceBaskets_2008-2020.xlsx) (дата обращения 01.06.2021)
48. Сведения об использовании информационных и коммуникационных технологий и производстве вычислительной техники, программного обеспечения и оказания услуг в этих сферах. № 3-информ. URL: <https://www.fedstat.ru/tools/storage/file?id=2106> (дата обращения 01.10.2021)

## DIGITAL INFRASTRUCTURE FOR BIG DATA

### Ershov, Peter Sergeevich

*Institute of the Information Society, head of Directorate of regional programs  
Moscow, Russian Federation  
peter.ershov@iis.ru*

### Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor  
Institute of the Information Society, chairman of the Board of directors  
Plekhanov Russian University of Economics, IIS-based Digital economy department, scientific advisor  
Moscow, Russian Federation  
yuri.hohlov@iis.ru*

### Abstract

*A conceptual framework and a set of indicators for monitoring and evaluating digital infrastructure for big data have been developed. The conceptual framework consists of three components: telecommunication infrastructure; computing infrastructure; platform infrastructure. Calculations of the main indicators for Russia and individual fields of activity are given; international comparisons are made as of the end of 2020. It was noted that Russia lags significantly behind the EU countries in terms of certain infrastructural indicators.*

### Keywords

*digital technology; big data; digital infrastructure; information infrastructure; telecommunication infrastructure; data transmission network; computing infrastructure; digital platform; monitoring and evaluation; Big Data for Digital Economy; BD4DE*

### References

1. TeleGeography's Global Bandwidth Research Service. URL: <https://www2.telegeography.com/global-bandwidth-research-service> (accessed 01.06.2021) Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (accessed 01.06.2021)
2. Global Industry Analysts. Internet Data Centers - Global Market Trajectory & Analytics. URL: <https://www.strategyr.com/market-report-internet-data-centers-forecasts-global-industry-analysts-inc.asp> (accessed 01.06.2021)
3. MarketsandMarkets. High Performance Computing (HPC) Market. URL: <https://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html> (accessed 01.06.2021)
4. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
5. Still, K., Seppänen, M., Korhonen, H., Valkokari, K., Suominen, A., & Kumpulainen, M. (2017). Business Model Innovation of Startups Developing Multisided Digital Platforms. In Y. Manolopoulos, M. Zaki, D. Diaz, & B. Theodoulidis (Eds.), 2017 IEEE 19th Conference on Business Informatics (CBI) (Vol. 2, pp. 70-75). [8012942] IEEE Institute of Electrical and Electronic Engineers. <https://doi.org/10.1109/CBI.2017.86>
6. Rossotto, C. M., Lal Das, P., Gasol Ramos, E., Clemente Miranda, E., Badran, M. F., Martinez Licetti, M., & Miralles Murciego, G. (2018). Digital platforms: A literature review and policy implications for development. *Competition and Regulation in Network Industries*, 19(1–2), 93–109. <https://doi.org/10.1177/1783591718809485>
7. Yaroslav Eferin, Yuri Hohlov, Carlo Rossotto (2019) «Digital platforms in Russia: competition between national and foreign multi-sided platforms stimulates growth and innovation», *Digital Policy, Regulation and Governance*, Vol. 21, Issue 2, pp.129-145. URL: <https://doi.org/10.1108/DPRG-11-2018-0065> (accessed 01.06.2021); Yeferin Ya.Yu, Rossotto K.M., Hohlov Yu.E. Tsifrovyye platformy v Rossii: konkurentsia mezhdu natsional'nymi i zarubezhnymi mnogostoronnimi platformami stimuliruyet ekonomicheskiy rost i

- innovatsii. Informatsionnoye obshchestvo, 2019, №1–2. С. 16–34. URL: <http://infosoc.iis.ru/article/view/135> (accessed 01.06.2021)
8. Jacobides, MG, Cennamo, C, Gawer, A. Towards a theory of ecosystems. *Strat Mgmt J.* 2018; 39: 2255–2276. <https://doi.org/10.1002/smj.2904>
  9. T.V. Ershova, Yu.E. Hohlov, S.B. Shaposhnik. Metodologiya monitoringa razvitiya i ispol'zovaniya tekhnologiy raboty s bol'shimi dannymi // *Informatsionnoye obshchestvo.* 2021. № 4–5. С. ??? <https://doi.org/10.1145/3453483.3454078>
  10. ITU. Measuring the Information Society Report. URL: <https://www.itu.int/pub/D-IND-ICTOI> (accessed 01.06.2021)
  11. World Telecommunication/ICT Indicators Database. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx> (accessed 01.06.2021)
  12. OECD Statistics by theme. URL: <https://stats.oecd.org/> (accessed 01.06.2021)
  13. Eurostat Statistics by theme. URL: <https://ec.europa.eu/eurostat/web/main/data/browse-statistics-by-theme> (accessed 01.06.2021)
  14. Федеральная служба государственной статистики. Официальная статистика. URL: <https://rosstat.gov.ru/folder/10705> (accessed 01.06.2021)
  15. OECD (2019), An Introduction to Online Platforms and Their Role in the Digital Transformation, OECD Publishing, Paris. <https://doi.org/10.1787/53e5f593-en>
  16. Analiz tekushchego sostoyaniya razvitiya tsifrovoy ekonomiki v Rossii. М.: Institut razvitiya informatsionnogo obshchestva, 2018. – 166 s. URL: <http://deca.iis.ru> (accessed 01.06.2021)
  17. Mossberger Karen, Tolbert Caroline J. Digital Citizenship and Digital Communities: How Technology Matters for Individuals and Communities. <https://doi.org/10.4018/IJEPR.20210701.0a2>
  18. Santos GL, Endo PT, Sadok D; Kelner J. When 5G Meets Deep Learning: A Systematic Review. URL: <https://doi.org/10.3390/a13090208>
  19. Seetharam Karthik, Kagiya Nobuyuki, Sengupta Partho P. Application of mobile health, telemedicine and artificial intelligence to echocardiography. <https://doi.org/10.1530/ERP-18-0081>
  20. Barricelli Barbara Rita; Casiraghi Elena; Fogli Daniela. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. <https://doi.org/10.1109/ACCESS.2019.2953499>
  21. Zacharov Igor, Arslanov Rinat, Gunin Maksim, Stefonishin Daniil, Bykov Andrey, Pavlov Sergey, Panarin Oleg, Maliutin Anton, Rykovanov Sergey, Fedorov Maxim. “Zhores” - Petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. <https://doi.org/10.1515/eng-2019-0059>
  22. Ved Mohit, Rizwanahmed B. Big Data Analytics in Telecommunication using state-of-the-art Big Data Framework in a Distributed Computing Environment: A Case Study. <https://doi.org/10.1109/COMPSAC.2019.00066>
  23. Zacharov I, Panarin O., Ryabinkin E., Izotov K., Teslyuk A. Virtualization for Scientific Workload. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000462194600017> ((accessed 01.06.2021)
  24. Liao Jianwei, Gerofi Balazs, Lien Guo-Yuan, Miyoshi Takemasa, Nishizawa Seiya, Tomita Hirofumi, Liao Wei-Keng, Choudhary Alok, Ishikawa Yutaka. A flexible I/O arbitration framework for netCDF-based big data processing workflows on high-end supercomputers. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000405351700005> (accessed 01.06.2021)
  25. Casero-Ripolles Andreu. Influencers in the Political Conversation on Twitter: Identifying Digital Authority with Big Data. <https://doi.org/10.3390/su13052851>
  26. Sampugnaro Rossana, Montemagno Francesca. In Search of the Americanization: Candidates and Political Campaigns in European General Election. <https://doi.org/10.1080/15377857.2020.1869832>
  27. Al-Abdallah Ghaith, Khair Nadine, Elmarakby Reyad. The Impact of Social Networking Sites on Luxury Vehicles Purchase Decision Process in Gulf Cooperation Council Countries. <https://doi.org/10.1080/08961530.2020.1867023>
  28. Fiofanova Olga Aleksandrovna. New literacy and data-future in education: advanced technology smart big-data. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000572957200014> (accessed 01.06.2021)

29. Yu Jun, Couldry Nick. Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. <https://doi.org/10.1080/1369118X.2020.1764604>
30. Efimov V, Lapteva A. Digital technology in higher education: situation analysis and prospects assessment (on the example of Krasnoyarsk Krai). <https://doi.org/10.21125/inted.2019.1392>
31. Larionova Viola, Sheka Andrey, Vasilyev Stanislaus. Students Behavioural Patterns on the National Open Education Platform. <https://doi.org/10.34190/EEL.19.126>
32. Schiavone Francesco, Leone Daniele, Sorrentino Annarita, Scaletti Alessandro. Re-designing the service experience in the value co-creation process: an exploratory study of a healthcare network. <https://doi.org/10.1108/BPMJ-11-2019-0475>
33. Pollock Daniel A. Digital Platforms as a Method of Invention for Infection Surveillance. <https://doi.org/10.1089/sur.2019.147>
34. Aledavood Talayeh, Hoyos Ana Maria Triana, Alakorkko Tuomas, Kaski Kimmo, Saramaki Jari, Isometsa Erkki, Darst Richard K. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. <https://doi.org/10.2196/resprot.6919>
35. Moller Arlen C., Merchant Gina, Conroy David E., West Robert, Hekler Eric, Kugler Kari C., Michie Susan. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. <https://doi.org/10.1007/s10865-016-9818-7>
36. Ivaschenko Anton, Korchivoy Stanislav, Spodobaev Michail. Infrastructural Models of Intermediary Service Providers in Digital Economy. [https://doi.org/10.1007/978-3-030-29513-4\\_44](https://doi.org/10.1007/978-3-030-29513-4_44)
37. Popov E.V. Econotronics of a smart city. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000476663600010> (accessed 01.06.2021)
38. Kiraly, Peter; Buechler, Marco. Measuring completeness as metadata quality metric in Europeana. URL: <https://www.webofscience.com/wos/woscc/full-record/WOS:000468499302103> (accessed 01.06.2021)
39. Ukaz Prezidenta Rossiyskoy Federatsii ot 21.07.2020 g. № 424 «O natsional'nykh tselyakh razvitiya Rossiyskoy Federatsii do 2030 goda». URL: <http://www.kremlin.ru/events/president/news/63728> (accessed 01.06.2021)
40. ITU. Measuring digital development ICT price trends 2020. URL: [https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU\\_ICTPriceTrends\\_2020.pdf](https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceTrends_2020.pdf) (accessed 01.06.2021)
41. TOP 500. URL: <https://www.top500.org/lists/top500/> (accessed 01.06.2021)
42. IDC. Lokomotivom razvitiya oblachnogo rynka stali platformnyye servisy. URL: <https://www.idc.com/getdoc.jsp?containerId=prEUR247478021> (accessed 01.06.2021)
43. Federal'naya sluzhba gosudarstvennoy statistiki. Itogi federal'nogo statisticheskogo nablyudeniya po forme № 3-inform «Svedeniya ob ispol'zovanii informatsionnykh i kommunikatsionnykh tekhnologiy i proizvodstve vychislitel'noy tekhniki, programmno obespecheniya i okazanii uslug v etikh sferakh». URL: <https://rosstat.gov.ru/folder/14478> (accessed 01.06.2021)
44. Eurostat database. URL: <https://ec.europa.eu/eurostat/web/main/data/database> (accessed 01.06.2021)
45. ITU Statistics. URL: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed 01.06.2021)
46. Rosstat. Itogi vyborochnogo federal'nogo statisticheskogo nablyudeniya po voprosam ispol'zovaniya naseleniyem informatsionnykh tekhnologiy i informatsionno-telekommunikatsionnykh setey. URL: [https://gks.ru/free\\_doc/new\\_site/business/it/ikt20/index.html](https://gks.ru/free_doc/new_site/business/it/ikt20/index.html) (accessed 01.06.2021)
47. ITU. The historical data for the price baskets for countries and regions. URL: [https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU\\_ICTPriceBaskets\\_2008-2020.xlsx](https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_ICTPriceBaskets_2008-2020.xlsx) (accessed 01.06.2021)
48. Svedeniya ob ispol'zovanii informatsionnykh i kommunikatsionnykh tekhnologiy i proizvodstve vychislitel'noy tekhniki, programmno obespecheniya i okazaniya uslug v etikh sferakh. № 3-inform. URL: <https://www.fedstat.ru/tools/storage/file?id=2106> (accessed 01.06.2021)