

Цифровая повестка, большие данные и официальная статистика

Георгий Константинович Оксенойт

Федеральная служба государственной статистики, г. Москва, Россия

В статье исследуются три группы вопросов: задачи, поставленные перед российской системой государственной статистики в рамках Программы «Цифровая экономика»; базовые характеристики больших данных с точки зрения организации статистического наблюдения; концепция умной статистики.

В качестве ключевого вопроса рассматривается понимание больших данных как «органических данных», являющихся данными первичного цифрового учета, осуществляемого в режиме реального времени программно-технологическими средствами без участия человека. В результате использования больших данных в официальной статистике формируется новый вид текущего (непрерывного) статистического наблюдения. Изначально цифровая форма существования больших данных и их базовые характеристики, в том числе в качестве микроданных, обуславливают как новые возможности измерения явлений и процессов, так и специфический характер ошибок наблюдения, возникающих при формировании больших данных. В целом организация наблюдения на основе больших данных рассматриваются автором как основной в перспективе вид статистического наблюдения.

В основе концепции умной статистики лежит непосредственное встраивание статистического наблюдения в систему первичного цифрового учета с последующей сквозной автоматизированной обработкой данных вплоть до получения агрегированных статистических показателей.

Рассмотренные подходы увязываются с задачами, поставленными перед российской системой государственной статистики в рамках Программы «Цифровая экономика Российской Федерации».

Ключевые слова: цифровая экономика, информационное общество, официальная статистика, большие данные, органические данные, статистическое наблюдение, ошибки наблюдения, качество больших данных, микроданные, наука о данных, умная статистика.

JEL: C81, C82, M15, O32.

Для цитирования: Оксенойт Г.К. Цифровая повестка, большие данные и официальная статистика. Вопросы статистики. 2018; 25(1):3-16.

Digital Agenda, Big Data and Official Statistics

Georgy K. Oksenoyt

Federal State Statistics Service, Moscow, Russia

This article reviews three groups of questions: objectives placed before the Russian official statistics by the Digital Economy program; basic characteristics of the Big Data in terms of data collection theory and design; concept of smart statistics.

The article addresses key issue of understanding Big Data as «organic data» that are primary digital records kept in real time by software and technology without human intervention. The use of Big Data in official statistics leads to development of the new type of continuous statistical observation. First and foremost, digital nature of Big Data and its basic characteristics, including as microdata, provide new possibilities for measuring phenomenon and processes, as well as determine the specific nature of observation errors that occur when Big Data are generated. In general, the organization and process of generating Big Data are viewed by the author as the main type of statistical observation in the future.

In the heart of the concept of smart statistics lays direct incorporation of statistical observation into the system of primary digital records followed by the straight through automatic data processing up to compiling statistical aggregates.

The approaches discussed in this article are linked to objectives placed before the Russian official statistics by the Digital Economy program.

Keywords: digital economy, information society, official statistics, Big Data, organic data, statistical observation, observation errors, quality of Big Data, microdata, data science, smart statistics.

JEL: C81, C82, M15, O32.

For citation: Oksenoyt G.K. Digital Agenda, Big Data and Official Statistics. *Voprosy statistiki*. 2018;25(1):3-16. (In Russ.)

Введение

Выступая на заключительной сессии 4-й Международной конференции по большим данным в официальной статистике в ноябре 2017 г. (Богота, Колумбия), трудно было предположить, что не пройдет и двух месяцев, как слова о том, что «будущие поколения статистиков будут работать преимущественно с большими данными», превратятся из футуристического прогноза в повседневную реальность уже для нынешнего поколения статистиков. Одновременно перед системой государственной статистики были поставлены практические задачи по реализации в кратчайшие исторические сроки, по своему существу, целого ряда принципов работы с большими данными в рамках создания новой технологической платформы, обеспечивающей удовлетворение информационных потребностей общества в условиях цифровой экономики.

В связи с этим в настоящей статье хотелось бы очертить круг задач, поставленных перед системой государственной статистики в этой области, остановиться на некоторых базовых характеристиках больших данных с точки зрения организации статистического наблюдения и кратко обозначить подходы к его модернизации в условиях цифровой экономики.

Все эти вопросы представляются важными для лучшего понимания перспектив использования больших данных в официальной статистике, а следовательно, для создания новой технологической платформы и построения на ее основе новой системы статистического наблюдения в условиях «новой статистической реальности».

Настоящая статья отражает личную позицию автора по рассматриваемым в ней вопросам.

Цифровая повестка

Полагаем, не будет большим преувеличением сказать, что 2017 г. прошел в России под знаком «цифровизации»¹. Начиная с конца 2016 г. был принят пакет стратегических документов и решений², направленных на создание условий для ускоренной цифровой трансформации во всех сферах жизнедеятельности российского общества.

В этих документах сформулированы основные цели и задачи, базовый понятийный аппарат, важнейшие направления планируемых преобразований, планы мероприятий (дорожные карты) и механизмы их реализации.

Российская государственная статистика, являясь одним из ключевых элементов информационного обеспечения системы принятия решений на всех уровнях управления обществом, государством и бизнесом, не может оставаться в стороне от данных процессов и должна занять в них соответствующее своему объективному назначению место.

В настоящее время в уже принятых документах перед Федеральной службой государственной статистики (в числе других федеральных органов исполнительной власти) поставлены следующие задачи:

1. В рамках нормативного и методического обеспечения мониторинга реализации Программы «Цифровая экономика Российской Федерации» (далее – Программы) осуществить: «разработку системы показателей мониторинга с учетом целей, задач и показателей настоящей Программы, рекомендаций международных организаций и иностранного опыта, а также инструментов мониторинга, включая модернизированные и новые формы федерального статис-

¹ Цифровизация – русскоязычный аналог английского (теперь уже ставшего международным) термина «datafication» [введен Кеннетом Кекьером и Виктором Майер-Шенбергером (Cukier K.N., Mayer-Schoenberger V. The Rise of Big Data // Foreign Affairs. May/June 2013) для обозначения процесса превращения всех явлений жизни в цифровую информацию] и его аналога – «digitalization».

² Указ Президента РФ от 05.12.2016 № 646 «Об утверждении Доктрины информационной безопасности Российской Федерации»; Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы»; Распоряжение Правительства РФ от 28.07.2017 № 1632-р «Об утверждении программы «Цифровая экономика Российской Федерации»; Постановление Правительства РФ от 28.08.2017 № 1030 «О системе управления реализацией программы «Цифровая экономика Российской Федерации» (вместе с «Правилами разработки, мониторинга и контроля выполнения планов мероприятий по реализации программы «Цифровая экономика Российской Федерации»); Решение Высшего Евразийского экономического совета от 11.10.2017 № 12 «Об основных направлениях реализации цифровой повестки Евразийского экономического союза до 2025 года»; Планы мероприятий по направлениям программы «Цифровая экономика Российской Федерации» [утв. Правительственной комиссией по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 18.12.2017 № 2)].

тического наблюдения за развитием цифровой инфраструктуры и цифровой трансформацией бизнеса, государственного и муниципального управления, образования, здравоохранения, использованием цифровых технологий населением и домохозяйствами, гармонизированные с аналогичными инструментами Организации экономического сотрудничества и развития; внесение изменений в федеральный план статистических работ, подготовку методических рекомендаций о порядке применения новых инструментов мониторинга»³.

Аналогичная задача сформулирована в рамках плана реализации «Стратегии развития информационного общества в Российской Федерации на 2017-2030 годы» (далее - Стратегии): осуществить «разработку статистического инструментария для оценки реализации настоящей Стратегии и мониторинга достижения значений показателей ее реализации»⁴.

2. В рамках разработки концепции среднесрочных мер по совершенствованию правового регулирования с целью развития цифровой экономики осуществить «определение новых правил сбора отчетности, в том числе статистической информации, исключающих дублирование этой информации, предусматривающих способы ее дистанционного получения и направленных на обеспечение потребностей общества и государства необходимыми данными в режиме реального времени»⁵, а также принять участие в формировании необходимых правовых условий для их внедрения и в подготовке и принятии соответствующих нормативных правовых актов⁶.

3. Представить предложения, направленные на совершенствование системы государственной статистики в части повышения оперативности получения отчетных данных и полноты собираемой информации в целях актуализации планов

мероприятий «Нормативное регулирование», «Формирование исследовательских компетенций и технологических заделов», «Информационная инфраструктура», «Информационная безопасность» и «Кадры и образование»⁷.

4. В рамках Плана мероприятий по направлению «Нормативное регулирование»⁸ Программы должны быть:

- проведена инвентаризация форм отчетности, в том числе при сборе статистической информации, разработка новых подходов к формированию отчетности [ожидаемый результат: создана карта данных, подлежащих предоставлению в государственные органы, выявлен перечень дублирующих и избыточных показателей; сформулированы рекомендации по оптимизации отчетности и исключению избыточного регулирования (в частности, определены подходы к переходу от сплошного статистического наблюдения к выборочному по отдельным формам отчетности); обеспечена возможность представления отчетности на порталы федеральных органов исполнительной власти с помощью программного доступа (API)] (мероприятие 01.01.009.001);

- осуществлена оценка целесообразности внедрения новых подходов к формированию отчетности, включая анализ перспектив развития горизонтального мониторинга и обмена данными [ожидаемый результат: сформулированы принципы реализации системы автоматического обмена данными, являющимися первичными данными для формирования отчетности, между участниками эксперимента и государственными органами, определены возможные варианты ценностных предложений для коммерческих компаний и компаний бюджетного сектора по автоматизации сбора статистической информации и иной отчетности, в том числе в режиме «реального времени»; ... сформирован перечень нормативных правовых актов, подлежащих изменению при переходе от

³ Программа «Цифровая экономика Российской Федерации», утв. распоряжением Правительства РФ от 28.07.2017 № 1632-р. Раздел V. Управление развитием цифровой экономики.

⁴ «Стратегия развития информационного общества в Российской Федерации на 2017-2030 годы». Раздел V. Пункт 54. Подпункт а).

⁵ Программа «Цифровая экономика Российской Федерации». «Дорожная карта». 1. Нормативное регулирование. Веха 1.3.1.

⁶ Программа «Цифровая экономика Российской Федерации». «Дорожная карта». 1. Нормативное регулирование. Вехи 1.13.1 и 1.13.2.

⁷ Из решений по итогам заседания Правительственной комиссии по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 18 декабря 2017 г. URL: <http://government.ru/orders/selection/401/30895>.

⁸ План мероприятий по направлению «Нормативное регулирование» программы «Цифровая экономика Российской Федерации» [утв. Правительственной комиссией по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 18.12.2017 № 2)].

отчетности (статистической, бухгалтерской, налоговой и др.) к реализации принципов автоматического обмена данными между предприятиями и государственными органами; ... проведена оценка целесообразности внедрения новых подходов к формированию отчетности] (мероприятие 01.01.009.002).

5. В рамках Плана мероприятий по направлению «Информационная инфраструктура»⁹ Программы должна быть создана цифровая аналитическая платформа в целях:

«предоставления данных, в том числе статистических, административных данных и нормативной справочной информации (НСИ), всем категориям пользователей;

загрузки респондентами различных видов отчетности и административных данных для использования органами государственной власти и местного самоуправления.

Ожидаемые результаты:

- разработана, введена в опытную, а затем в промышленную эксплуатацию единая цифровая аналитическая платформа предоставления статистических, административных данных и НСИ, которая обеспечивает для всех категорий пользователей:

однократность предоставления данных, в том числе статистических и административных, во все органы государственной власти всех уровней и местного самоуправления, в том числе в онлайн-режиме;

возможность формирования и использования аналитических показателей для целей государственного управления (включая вопросы формирования традиционных статистических показателей) с использованием доступной базы первичных статистических данных в соответствии с меняющимися информационными потребностями;

предоставление всех данных и НСИ (респондентами и по запросам пользователей) исключительно в электронной форме;

одновременное использование данных различной природы [данных статистических переписей и обследований, административных данных (ГИС, реестров, регистров) и альтернативных источников информации], что позволит опера-

тивно управлять бизнес-процессами с учетом всей полноты информации;

интеграцию бухгалтерской, статистической и налоговой отчетности экономических агентов;

- цифровая аналитическая платформа предоставления статистических, административных данных и НСИ интегрирована с другими платформами и обеспечивает возможность загрузки респондентами различных видов отчетности и административных данных для использования органами государственной власти и местного самоуправления» (мероприятие 04.03.015.004)¹⁰.

Поставленные задачи отражают существующий общественный запрос в отношении выполнения системой государственной статистики ее важнейших функций в условиях формирования цифровой экономики.

В данной статье было бы нецелесообразно останавливаться на задачах, связанных с измерением цифровой экономики. Это большая самостоятельная тема, над которой в мировом статистическом сообществе и у нас в России уже не первый год идет интенсивная работа, и которая лишь получила дополнительный стимул в связи с принятием перечисленных выше решений.

Общие подходы к решению комплекса задач, связанных с оптимизацией системы статистической отчетности, ведением нормативно-справочной информации, устранением дублирования, снижением нагрузки на респондентов и ряд других, были намечены в докладе «Целевая модель «Статистика - 2025» [1] в рамках концепции «одного окна» для сбора и распространения статистической информации и единого регистра форм и показателей статистической, административной отчетности и нормативно-справочной информации и также отдельно рассматриваться не будут.

В условиях цифровой трансформации решение всех вопросов развития системы государственной статистики должно быть тесным образом увязано с достижением общей цели Программы - созданием «экосистемы цифровой экономики Российской Федерации, в которой данные в цифровой форме являются ключевым фактором производства во всех сферах социально-экономической деятельности и в которой обеспечено

⁹ План мероприятий по направлению «Информационная инфраструктура» программы «Цифровая экономика Российской Федерации» [утв. Правительственной комиссией по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 18.12.2017 № 2)].

¹⁰ А также определена «модель коммерциализации» цифровой аналитической платформы (мероприятие 04.03.015.004.002 Программы).

эффективное взаимодействие, включая трансграничное, бизнеса, научно-образовательного сообщества, государства и граждан»¹¹, что требует качественно новых подходов практически во всех сферах статистической деятельности.

При выработке новых перспективных подходов государственная статистика должна опираться на «основные сквозные технологии» (разумеется, по мере их развития), которые входят в рамки Программы «Цифровая экономика Российской Федерации»: «большие данные; нейротехнологии и искусственный интеллект; системы распределенного реестра; квантовые технологии; новые производственные технологии; промышленный интернет; компоненты робототехники и сенсорики; технологии беспроводной связи; технологии виртуальной и дополненной реальностей»¹², а также на весь комплекс цифровых инфраструктурных и иных преобразований, намеченных в Стратегии развития информационного общества и в Программе. Только в этом случае государственная статистика сможет стать «органичным» элементом создаваемой экосистемы цифровой экономики и обеспечить полноценное выполнение возложенных на нее функций в рамках формирующегося информационного общества.

Цифровая трансформация экономики и общества должна стать для государственной статистики не только объектом статистического наблюдения; система государственной статистики должна сама подвергнуться всесторонней цифровой трансформации. При этом не только с точки зрения общего процесса перевода на новые платформы и технологии органов государственной власти и управления, но и активного формирования заказа на архитектуру и функционал создаваемых платформ и технологий, их интеграции, дальнейшего эффективного развития.

Исходя из вышеизложенного, представляется целесообразным коротко остановиться на ряде ключевых вопросов, связанных с предстоящей цифровой трансформацией экономики, общества и самой системы государственной статистики.

Большие данные

Большие данные не случайно выступают первыми в перечне «основных сквозных технологий»,

с самого начала включенных в Программу «Цифровая экономика Российской Федерации». Революция данных является основным компонентом процесса цифровой трансформации общества, поэтому от того или иного решения вопроса об отношении официальной статистики к большим данным во многом зависит ее дальнейшая судьба как в среднесрочной, так и долгосрочной перспективе.

Вопрос использования в официальной статистике больших данных уже более 10 лет находится в поле зрения мирового статистического сообщества, крупнейших международных организаций (ОЭСР, МВФ, Всемирного банка и др.); с 2013 г. вошел в регулярную повестку ЕЭК ООН и Европейской статистической системы, а с 2014 г. – Статистической комиссии ООН. Этой теме во всем мире посвящено огромное количество официальных публикаций, исследовательских и прикладных работ, в том числе целый ряд материалов публиковался на страницах журнала «Вопросы статистики», включая последний международный обзор в декабрьском выпуске 2017 г. [2-4].

Общий итог на сегодняшний день выглядит следующим образом.

Официальная статистика на всех уровнях (от национальных статистических служб до международных организаций) в целом придерживается консервативной позиции, с одной стороны, подчеркивая высокую потенциальную значимость больших данных, а с другой – постоянно напоминая о рисках и ограничениях в использовании больших данных в официальной статистике.

Количество находящихся на разных стадиях реализации пилотных проектов по использованию больших данных для формирования официальной статистической информации непрерывно растет, хотя и остается сравнительно небольшим, а сферы их применения – относительно узкими. В настоящее время это – преимущественно проекты по использованию интернет-ресурсов и результатов сканирования в статистике потребительских цен, данные мобильных операторов в статистике туризма, миграции и мобильности населения, данные спутниковой съемки в статистике сельского хозяйства и окружающей среды, данные дорожных датчиков в транспортной статистике, данные социальных сетей в статистике рыночной конъюнктуры и некоторые другие.

¹¹ Программа «Цифровая экономика Российской Федерации». Раздел I. Общие положения.

¹² Там же.

Одновременно в большинстве ведущих стран (в национальных статистических службах) и в международных организациях идет энергичная практическая работа в области больших данных: создаются специализированные подразделения; разрабатываются стратегии и дорожные карты; аллоцируются финансовые ресурсы; ведутся активные экспериментальные работы; осуществляется разработка статистической методологии, математического аппарата, программного обеспечения.

Помимо этого, формируются специальные (в том числе открытые) национальные и международные технологические платформы для проведения экспериментов; на основе кооперации с ведущими академическими и учебными институтами и бизнес-структурами создаются национальные и международные центры по подготовке кадров; формируется необходимая законодательная и нормативная правовая база; публикуются различные материалы (исследовательские, результаты экспериментальных расчетов, методические рекомендации, обзоры лучших практик); идет активный международный обмен опытом.

Серьезный анализ в данной области [5] был подготовлен группой специалистов Международного валютного фонда [6].

В целом можно констатировать, что превалирует прагматический подход, предполагающий активные практические действия, в том числе экспериментального характера (разумеется, с учетом имеющихся рисков и ограничений), в то время как общие методологические, организационные, правовые и иные проблемы решаются по мере накопления соответствующего опыта.

В российской государственной статистике практические работы по использованию больших данных до настоящего времени не велись. Их начало планируется в 2018 г. в области использования больших данных в статистике потребительских цен, статистике населения и ряде других отраслей. Наряду с этим, в 2018 г. планируется разработка общей концепции использования больших данных в официальной статистике. Для этого, как отмечалось выше, имеются серьезные методологические и практические заделы на международном уровне.

В настоящее время наиболее разработанными в методологическом плане являются вопросы, связанные с качеством больших данных, рисками

и ограничениями их использования в официальной статистике: слабая структурированность, нестабильность и непрозрачность источников, неполный охват, непонятное качество, проблемы конфиденциальности и др. (пакет рекомендаций в этой области был разработан ЕЭК ООН [7]).

Это является вполне закономерной реакцией системы государственной статистики (в международном масштабе). Она обусловлена, во-первых, исторически сложившимся высоким уровнем профессионализма и ответственности официальной статистики перед обществом; во-вторых, стремлением защитить свои традиционные ценности как общественного блага – полноту, достоверность, объективность, научную обоснованность, сопоставимость, конфиденциальность и др.

Кроме того, это – понятное желание обеспечить для себя равные конкурентные условия на рынке информационных услуг в сравнении с частными корпорациями, являющимися основными владельцами больших данных и поставщиками альтернативных статистических оценок. Частные компании в борьбе за информационные рынки склонны жертвовать качеством в обмен на скорость предоставляемой ими информации. Нужно отметить, что антитеза «качество или скорость» вообще является одной из основных дилемм развития современной информационной индустрии, включая систему государственной статистики, и выбор здесь совсем не очевиден, так как потребитель – он же заказчик статистической информации всегда будет стремиться к оптимальному соотношению скорости, цены и качества, часто отдавая предпочтение первым двум.

Достаточно полно описаны также общеизвестные основные достоинства больших данных: высокая скорость обновления, возможность оперативного сбора, обработки и предоставления, многообразие охватываемых явлений, детальность описания объектов наблюдения, снижение нагрузки на респондентов, относительно низкая себестоимость и др., а также ряд практических вопросов, требующих серьезных дополнительных усилий: укрепление кадрового потенциала, формирование новых навыков и компетенций, создание государственно-частных партнерств, большие вычислительные мощности, принципиально новые технологии сбора, хранения и обработки данных, обеспечение конфиденциальности и др.

В связи с этим, не предвзято результатов запланированных работ (и не вступая в полемику), хотелось бы кратко остановиться на ряде методологических вопросов, в понимании которых, на наш взгляд, до сих пор отсутствует определенность и имеются некоторые пробелы, затрудняющие объективную оценку возможностей и перспектив использования больших данных в официальной статистике.

Прежде всего, это относится к пониманию базовых с точки зрения статистики характеристик больших данных.

Как отмечалось выше, в целом в указанной области до настоящего времени преобладает сугубо практический, ориентированный на проведение экспериментов подход. Это проявляется, например, в преимущественно описательном характере классификации больших данных, которая была разработана ЕЭК ООН [8], а затем модифицирована Европейской статистической комиссией [9]. В обоих случаях большие данные с относительно небольшими различиями классифицируются по источникам их происхождения, во втором случае – с привязкой к отраслям статистики, в которых они могут быть преимущественно использованы для проведения пилотных исследований, а не с точки зрения их основных внутренних характеристик.

То же относится к большинству определений больших данных, основное внимание в которых, как правило, сосредоточивается на их важнейших и наиболее очевидных свойствах, таких, как большой объем, большая скорость обновления, большое разнообразие, высокая изменчивость и др.

При этом в центре внимания оказываются свойства больших данных как больших *совокупностей* данных, наиболее важные с практической точки зрения их обработки и анализа, а не свойства больших данных как количественных характеристик явлений как таковых.

Вместе с тем все свойства больших данных тесно связаны между собой, и только их совместное рассмотрение может задать правильный вектор их дальнейшего использования в официальной статистике.

В связи с этим, как нам представляется, принципиально важно рассматривать не только пе-

речисленные выше свойства больших данных как больших совокупностей данных, но и следующие, изначально присущие большим данным свойства. Большие данные:

1. Образуются непосредственно в ходе протекания социально-экономических и иных явлений и процессов (как побочные продукты этих процессов) в качестве их цифрового образа;

2. Изначально и в течение всего жизненного цикла существуют исключительно в цифровой форме;

3. Представляют собой либо непосредственную (и единственную) форму существования цифровых явлений, либо образуются в результате непосредственного наблюдения и регистрации в цифровой форме иных явлений¹³;

4. Процессы наблюдения, формирования цифровых образов, регистрации и документирования больших данных совпадают¹⁴. При этом признание факта документирования больших данных в момент их регистрации в электронной форме зависит от национального законодательства (в России в настоящее время не признается);

5. Наблюдение осуществляется автоматическими программно-технологическими средствами (средствами объективного наблюдения) без участия человека.

Вследствие перечисленных выше свойств большие данные обладают рядом чрезвычайно важных с точки зрения организации статистического наблюдения позитивных характеристик:

- формируются в режиме реального времени
- отсутствует разрыв во времени между совершением события и его регистрацией; при этом регистрация осуществляется непрерывно в течение установленного периода наблюдения (с учетом объективных возможностей программно-технологических средств наблюдения, свойств объектов наблюдения, объективных условий ведения и заданных параметров наблюдения);

- становятся доступными для последующей обработки (в статистических или иных целях) в режиме реального времени с момента их регистрации;

- обладают высокой степенью полноты, так как выступают результатом сплошного наблю-

¹³ Здесь и далее в соответствии с целями настоящей статьи под «цифровыми» явлениями будут пониматься явления и процессы, изначально возникающие и существующие исключительно в цифровой форме (например, сообщения в социальных сетях, данные мобильных операторов, информация о ценах интернет-магазинов), в отличие от «иных» явлений, в отношении которых могут формироваться их цифровые образы, также выступающие в качестве больших данных (например, спутниковые изображения, данные датчиков дорожного движения, размещаемые в Интернете цены обычных магазинов).

¹⁴ Далее все эти действия вместе будем называть «наблюдение».

дения явлений и осуществляются непрерывно (например, каждая камера в течение периода своей работы фиксирует изображения всех движущихся участников дорожного движения на участке своего наблюдения с учетом условий видимости и своих технических возможностей; данные мобильных операторов отражают всю совокупность событий в своих цифровых сетях - все включения-выключения всех мобильных устройств, все их перемещения в пространстве и продолжительность работы во времени, все входящие и исходящие сигналы, их характер, продолжительность и направление и т. п.)¹⁵;

- могут быть достаточно четко структурированы (за исключением больших данных, относящихся по классификации ЕЭК ООН к классу «генерируемых человеком» [8]);

- обладают высоким уровнем достоверности, так как не содержат ошибок наблюдения, обусловленных «человеческим фактором» (за исключением ошибок, относящихся к «преднамеренным», то есть целенаправленному «сокрытию информации, фальсификации»¹⁶).

Описанный выше подход соответствует пониманию больших данных как «органических данных» («organic data»), возникающих непосредственно в процессе функционирования объектов наблюдения в качестве их побочных продуктов («by-products»), в отличие от «сконструированных данных» («designed data»), возникающих в результате специально организованных статистических наблюдений [2, 6, 11-13].

В целом большие данные представляют собой всеобщую цифровую форму существования (проявления) социально-экономических и иных явлений и процессов. При этом с точки зрения статистики, большие данные по своей сути являются данными первичного учета¹⁷ (при условной применимости этого понятия к изначально циф-

ровым явлениям). В результате должен сформироваться еще один вид текущего (непрерывного)¹⁸ статистического наблюдения - на основе больших данных как данных всеобщего первичного (цифрового) учета, вносящий также новации в понимание форм и способов статистического наблюдения¹⁹.

С точки зрения исследовательских целей, большие данные представляют собой максимально детализированные микроданные о наблюдаемых явлениях и процессах, изменяющиеся в режиме реального времени. Как таковые, большие данные впервые позволяют преодолеть противоречие между детализированностью статистической информации и скоростью ее актуализации, делая получение микроданных более оперативным, чем получение основанных на них агрегированных показателей.

Рассмотренные выше базовые характеристики больших данных позволяют по-новому взглянуть на причины их ограниченного использования в официальной статистике. Как представляется, эти причины кроются не во «врожденных» (и, следовательно, неустранимых) недостатках больших данных, а в той объективной стадии цифровой трансформации, на которой мы находимся, выражающейся в недостаточном уровне развития цифровой экономики, информационной инфраструктуры и цифровой культуры общества.

Ускоренное развитие всех трансформационных процессов предусмотрено в перечисленных в начале данной статьи документах, и, по мере их реализации, будет объективно улучшаться качество больших данных и расширяться сфера их применения в официальной статистике, а именно:

- неизменно повышаться точность, скорость и качество работы программно-технологических средств наблюдения (точность геопозициониро-

¹⁵ При этом необходимо помнить о возможных различиях в объектах наблюдения. В случае с мобильными операторами объектом наблюдения является совокупность SIM-карт, а для статистики населения или туризма - совокупность людей.

¹⁶ Здесь и далее типы ошибок наблюдения приводятся в классификации О. Моргенштерна, использованной в его фундаментальной и до настоящего времени сохраняющей свою актуальность монографии «О точности экономико-статистических наблюдений» [10].

¹⁷ В ряде случаев первичный учет называют также оперативным или оперативно-техническим учетом.

¹⁸ *Текущее наблюдение* - это наблюдение, при котором факты регистрируются по мере их возникновения, то есть учитываются постоянно, непрерывно (классический пример - регистрация рождения, смерти в актах гражданского состояния). Здесь и далее виды, формы и способы статистического наблюдения будут рассматриваться в соответствии с их пониманием в «Энциклопедии статистических терминов» (статья «Наблюдение статистическое». Том 1 «Методологические основы статистики». М.: Федеральная служба государственной статистики, 2013. URL: http://www.gks.ru/free_doc/new_site/rosstat/stbook11/tom1.pdf).

¹⁹ Так, с точки зрения: формы статистического наблюдения, большие данные в качестве данных первичного учета не являются ни отчетностью, ни специальными обследованиями; периодичности - текущее наблюдение в режиме реального времени; охвата - сплошное наблюдение на микроуровне; способов статистического наблюдения - модификацию непосредственного наблюдения на основе саморегистрации, осуществляемой программно-технологическими средствами без участия человека.

вания, качество изображения, скорость обновления и т. п.), а следовательно, точность, скорость и качество формирования цифровых образов (больших данных);

- расти круг охватываемых большими данными явлений и процессов (охват населения Интернетом и мобильной связью, введение уникальных идентификаторов личности, использование безналичных расчетов, покрытие сетью датчиков дорожного движения и спутниковыми изображениями, развитие Интернета вещей, включая индустриальный Интернет, и т. д. и т. п.) и, следовательно, постепенно приближаться и, в конечном счете, во многих случаях совпадать с генеральной совокупностью наблюдаемых явлений в официальной статистике;

- осуществляться переход на новые технологии (в том числе на технологии распределенных регистров или «block-chain»), гармонизация и интеграция (между собой и с государственными информационными системами) всех видов административных регистров, часто рассматриваемых в качестве одной из разновидностей больших данных, несмотря на наличие во многих случаях «человеческого фактора» при их формировании;

- активно развиваться гармонизация и стандартизация структуры различных источников больших данных, осуществляемая для удобства хозяйственного оборота и приводящая к ее большей сопоставимости (в пределе - идентичности) со структурой метаданных официальной статистики на национальном и международном уровнях.

Наряду с перечисленными выше объективными процессами цифровой трансформации, обуславливающими повышение роли больших данных, должна также развиваться методология работы с большими данными официальной статистики.

На сегодняшний день это развитие идет во всем мире чрезвычайно высокими темпами в рамках нового междисциплинарного направления «науки о данных» или «data science». При этом активно развиваются новые технологии сбора, хранения и обработки больших данных, сами же концепции статистического измерения тех или иных явлений за редкими исключениями, примеры которых приведены в работе [6], остаются неизменными, такими, какими они сформировались в условиях «доцифрового» («доинтернетного») уклада экономики и социального образа жизни общества или

на их относительно ранних стадиях развития (а иногда еще в «докомпьютерную» эпоху).

В связи с этим большие данные часто рассматриваются официальной статистикой с позиции традиционных технологий как совокупность данных, на которых надо построить статистическое наблюдение (выборку) с целью извлечения результата, структурированного в соответствии с ранее разработанными концепциями измерения соответствующих явлений, либо изначально аналогичным образом структурировать большие данные.

В ряде случаев это возможно, но, как правило, достаточно трудоемко (например, в статистике цен). Задача изначальной (при формировании) стандартизации и структурирования больших данных, очень важна особенно, для «процессно-генерируемых» больших данных [6, 8], но является относительно ресурсоемкой для бизнеса.

Вместе с тем представляется, что в результате развития цифровой экономики и информационного общества происходит как изменение сущности и форм проявления многих процессов и явлений, так и объективные изменения в возможностях их измерения. Например:

- процесс изменения качества большинства «цифровых» товаров и услуг становится практически непрерывным, что ставит вопрос о целесообразности сохранения классического требования об элиминировании изменений их качества при расчете индексов потребительских цен;

- есть существенная разница в требованиях к точности отдельных наблюдений (или используемых в расчетах весов) при наличии десятков или сотен наблюдений и при наличии сотен тысяч или миллионов наблюдений;

- появляется возможность регулярного наблюдения не только за массовыми, но и за «малыми» явлениями (отклонения, уникальные, в том числе аномальные явления, новые явления, находящиеся на ранних стадиях развития)²⁰.

Отметим, что разработка методологии выявления и оценки ошибок наблюдения в отношении больших данных также, безусловно, является одной из важных задач официальной статистики, особенно с учетом того, что цифровая природа больших данных определяет существенные особенности связанных с ними ошибок наблюдения.

Так, упомянутые ранее «преднамеренные» ошибки в отношении больших данных приобре-

²⁰ До настоящего времени такие «выбросы» данных, как правило, рассматривались как ошибки наблюдения и либо сглаживались, либо просто исключались из расчетов.

тают форму умышленного создания цифровых образов несуществующих явлений или искажения цифровых образов существующих явлений (например, генерируемые роботами «спам», массовые поисковые запросы в Интернете или сообщения в социальных сетях). В предельном случае – препятствования созданию цифрового образа явления в целом (например, запрещенные к использованию на транспортных средствах так называемые «антирадары»).

Новыми, специфическими для больших данных типами ошибок наблюдения являются, например:

- продолжение самостоятельного существования в информационном пространстве (Интернет) ранее сформированных цифровых образов явлений после того, как сами явления уже прекратили свое существование (задержка обновления информации на сайтах, «брошенные» сайты);

- создание и самостоятельное существование в информационном пространстве (Интернет) множества копий (полных или модифицированных) цифрового образа одного и того же явления («зеркальные сайты») либо различных цифровых образов одного и того же объекта (реклама), по своему эффекту для статистики аналогичные ошибкам «двойного учета», но с потенциально более высоким мультиплицирующим эффектом.

Данные ошибки широко распространены, часто трудно идентифицируемы и требуют разработки новых методов (и технологий) их выявления, оценки и элиминирования.

Принципиально новым моментом с точки зрения теории и практики организации статистического наблюдения является наличие угрозы возникновения специфических форс-мажорных обстоятельств. Эти обстоятельства делают невозможным в течение неопределенного периода времени формирование больших данных (цифровых образов тех или иных явлений) либо приводят к их массовому искажению в результате кибер (хакерских) атак или под влиянием техногенных факторов (например, нарушений энергоснабжения). Вследствие воздействия указанных факторов возможны также массовые нарушения конфиденциальности. В связи с этим особую роль в данной области играют вопросы обеспечения информационной безопасности.

В целом расширяющееся использование больших данных будет означать для официальной

статистики постепенный переход от парадигмы периодических выборочных наблюдений в сочетании с единовременными сплошными обследованиями (переписями) к парадигме непрерывных (в режиме реального времени) сплошных наблюдений на основе больших данных (включая административные источники) в сочетании с периодическими выборочными обследованиями, сопровождающийся сокращением (а в перспективе – полным отмиранием) статистической отчетности как формы наблюдения.

При этом будет осуществляться постепенный переход от использования больших данных как дополнительного источника информации для верификации данных официальной статистики к использованию данных выборочных статистических наблюдений для верификации официальной статистической информации, сформированной на основе больших данных.

С точки зрения методологии статистических исследований будет происходить кардинальное расширение сферы использования методов статистического анализа, моделирования и прогнозирования на основе больших данных как микроданных, изменяющихся в режиме реального времени.

Это создает принципиально новые возможности по сравнению с данными статистической отчетности или единовременных обследований, в том числе на макроуровне – в части создания системы опережающих и косвенных индикаторов экономического и социального развития, на микроуровне – в части индивидуальных и групповых моделей поведения экономических субъектов и населения.

Необходимо также отметить, что в силу своей природы большие данные впервые создают возможность для непосредственного наблюдения за латентными (теневыми) явлениями и процессами в экономике и социальной сфере, которые практически не улавливаются традиционными формами статистического наблюдения.

Умная статистика

Выше мы рассмотрели некоторые, ключевые на наш взгляд, характеристики больших данных с точки зрения организации статистического наблюдения, понимания перспектив и последствий их использования в официальной статистике. Однако такое рассмотрение было бы неполным,

если бы мы не включили в поле нашего зрения, наряду с вопросами методологии, также организационно-технологические вопросы модернизации системы статистического наблюдения в эпоху больших данных, которые в первой части настоящей статьи были обозначены как задача цифровой трансформации системы государственной статистики.

В целом эта задача может быть сформулирована как задача построения системы «smart statistics», или «умной статистики» (по аналогии с умным городом, умным домом и др.). Концепция умной статистики была сформулирована в рамках Рабочей группы по большим данным Евростата и опубликована в качестве рабочего материала в феврале 2017 г. [14].

Необходимо отметить, что еще в 1999-2000 гг. аналогичные идеи развивались в российской статистике (на региональном уровне) в рамках концепции «электронной статистики». Информация об этих работах была опубликована И.В. Суловым в упоминавшейся выше статье на страницах журнала «Вопросы статистики»[2].

Ключевая идея концепции умной статистики (далее - умная статистика или умная система статистики) состоит в сквозной автоматизации всего процесса производства статистической информации и его «органическом» (на программно-технологическом уровне) встраивании в также полностью автоматизированный процесс первичного цифрового учета.

Эта идея самым тесным образом связана с рассмотренным нами выше пониманием больших данных как «органических данных» и по своему существу сводится к изначальному «органическому» (программно-технологическому) встраиванию статистического наблюдения непосредственно в процесс формирования больших данных, по сути приводящему к объединению, а в ряде случаев - к слиянию, в единый, полностью автоматизированный технологический процесс двух в настоящее время существующих независимо друг от друга процессов - первичного цифрового учета (формирования больших данных) и статистического наблюдения (формирования показателей статистической отчетности)²¹.

Разумеется, все последующие этапы процесса производства статистической информации в соответствии с концепцией умной статистики также должны быть полностью автоматизированы²².

Концепция «органического» (на программно-технологическом уровне) объединения (слияния) процессов формирования данных первичного учета и формирования показателей статистической отчетности является принципиально важной по следующим основным причинам:

- именно на стадиях формирования данных первичного учета (сфера больших данных) и формирования на их основе первичных статистических данных - данных первичной статистической отчетности (сфера статистического наблюдения) возникают основные, наиболее частые, масштабные и трудно устранимые ошибки наблюдения, в первую очередь связанные с «человеческим фактором», включая «преднамеренные» ошибки и др.;

- именно эти две стадии являются наиболее трудоемкими, сложно поддающимися автоматизации (особенно стадия первичного учета) и, следовательно, создающими основную отчетную нагрузку на респондентов с точки зрения трудовых, временных и финансовых затрат (стадия предоставления уже сформированной статистической отчетности в органы государственной власти является существенно менее трудоемкой и дорогостоящей, чем стадии ведения первичного учета и составления отчетности);

- именно полная автоматизация процессов формирования данных первичного учета, статистической и иной отчетности позволит в наибольшей степени снизить отчетную нагрузку на респондентов, повысить качество, детальность разработки и оперативность предоставления статистической информации.

Вместе с тем необходимо отметить, что концепция умной статистики требует наличия «на другой стороне», в качестве своего «зеркального» отражения, концепции умной фабрики, умного магазина, умной больницы и т. п. При этом полная интеграция двух умных систем всегда является очень тонким вопросом, не столько с точки зрения программно-технологических решений

²¹ Здесь и далее мы будем говорить о статистической отчетности и первичном (цифровом) учете, имея в виду, что рассматриваемые вопросы в равной мере относятся ко всем видам отчетности и первичного учета.

²² Аналогичные технологии «сквозной автоматической обработки данных» («straight through processing» или STP) известны в финансовом мире уже более 20 лет и в настоящее время являются общепринятым (обязательным) стандартом осуществления клиринга и расчетов на всех мировых организованных финансовых рынках, включая российский.

и необходимости полномасштабного развития Интернета вещей и индустриального Интернета, сколько в связи с возникновением большого количества правовых и иных проблем, связанных, прозрачностью, конфиденциальностью, информационной безопасностью и др.

В целом комплекс задач по построению умной системы статистики (без использования данного термина), за исключением задачи автоматизации первичного учета, практически полностью сформулирован в рамках Программы «Цифровая экономика Российской Федерации» и планов мероприятий по направлениям «Информационная инфраструктура» и «Нормативное регулирование».

Разумеется, повсеместное внедрение «идеальной модели» умной системы статистики в силу объективных причин потребует достаточно длительного периода времени (уровень развития Интернет-технологий, состояние и скорость обновления основных фондов, традиции ведения бизнеса, наличие теневого сектора, необходимые инвестиции, информационная культура населения, правовые проблемы и др.), но в отношении высокотехнологичных компаний цифрового сегмента экономики и в первую очередь его государственного сектора, представляется вполне реализуемым в ближайшем будущем.

Заключение

Необходимость работы в условиях ускоренного развития цифровой экономики, усиливающейся конкуренции на рынке информационных услуг и экспансии больших данных на территорию официальной статистики, поставленные задачи по цифровой трансформации, развитию нормативного регулирования и информационной инфраструктуры требуют от российской государственной статистики активных практических действий в указанном направлении.

Большая часть необходимых мер зафиксирована в рассмотренных правительственных документах, в первую очередь в рамках задачи по созданию «цифровой аналитической платформы для сбора и предоставления данных, в том числе статистических, административных данных и нормативно-справочной информации, всем категориям пользователей»²³ (далее - ЦАП). Это:

- обеспечение однократности предоставления респондентами всех видов собираемых государством данных;

- предоставление всех данных и НСИ (респондентами и по запросам пользователей) исключительно в электронной форме;

- обеспечение автоматического обмена данными между респондентами и ЦАП, в том числе в режиме реального времени;

- интеграция ЦАП с другими государственными информационными системами и обеспечение между ними автоматического обмена данными, в том числе в режиме реального времени;

- интеграция бухгалтерской, статистической и налоговой отчетности и др.

С точки зрения создания умной системы статистики, основное, что необходимо добавить к перечисленным выше мерам (с учетом существенно более длительного периода внедрения), это обеспечение на программно-технологическом уровне интеграции бухгалтерской, статистической и налоговой отчетности с системой первичного (оперативного) учета и автоматизация самой системы первичного (оперативного) учета.

Все эти задачи, безусловно, должны быть поэтапно реализованы, однако для успешной реализации необходимо также создание определенных организационно-правовых условий.

Основные шаги, направленные на создание таких условий (на том этапе - в части больших данных), были сформулированы в работах [1, 3]. Это прежде всего:

- разработка и принятие комплекса законодательных и нормативно-правовых актов, обеспечивающих возможность использования больших данных в официальной статистике и стимулирующих автоматизацию и стандартизацию первичного (оперативного) учета и его интеграцию с системой бухгалтерской, статистической и налоговой отчетности, внедрение процессов сквозной автоматической обработки данных;

- формирование с участием органов государственной статистики и других заинтересованных сторон системы частно-государственных партнерств (в широком смысле), центров компетенции, координационных центров по использованию больших данных в официальной статистике, в том числе выступающих «удостоверяющими

²³ План мероприятий по направлению «Информационная инфраструктура» программы «Цифровая экономика Российской Федерации».

центрами» в отношении качества больших данных и обеспечения их конфиденциальности²⁴;

- создание (возможно, в рамках ЦАП) единого ресурсного центра, включая открытую платформу для проведения экспериментов по обработке больших данных в целях их использования в официальной статистике;

- разработка системы подготовки и повышения квалификации кадров для работы с большими данными;

- развитие международного сотрудничества в рамках Евразийского экономического союза и иных международных организаций.

Дополнительно хотелось бы отметить, что существенно более высокая скорость изменений, краткость жизненного цикла явлений в цифровой экономике требуют кардинального ускорения процессов разработки и внедрения новой статистической методологии, формирования качественно новых подходов к измерению цифровых явлений и процессов, проведения экспериментов по использованию больших данных в официальной статистике, внедрению технологий сквозной автоматизированной обработки данных (в рамках создания ЦАП).

Выработка новых подходов невозможна без проведения экспериментальных работ. Это право и соответствующий функционал должны быть закреплены за органами государственной статистики на нормативном правовом уровне, в том числе представляется возможным использование в этих целях предусмотренного Программой «Цифровая экономика Российской Федерации» механизма «регуляторных песочниц».

В заключение хотелось бы отметить еще один важный момент: «Ситуация меняется. Планы мероприятий, которые мы рассматриваем сегодня, завтра уже могут стать менее актуальными, а послезавтра просто устареть. Здесь среда развивается быстрее, чем наше регулирование. Чтобы этого не допустить, сформирована гибкая система управления изменениями ... используется единая платформа информационного взаимодействия. Надеюсь, что она позволит избежать капитальных ошибок»²⁵.

В 2018 г. предстоит интенсивная работа по реализации поставленных задач, а в дальнейшем жизнь в условиях «новой статистической реальности».

Литература

1. **Оксенойт Г.К.** Целевая модель «Статистика-2025». Доклад на заседании Совета по открытым данным Правительственной комиссии по координации деятельности открытого правительства от 23 декабря 2016 г. URL: http://www.gks.ru/free_doc/new_site/rosstat/smi/reлиз-sod.pdf.

2. **Суслов В.И.** Об экономических измерениях: вероятность и достоверность, математическое моделирование, большие данные, электронная статистика // Вопросы статистики. 2016. № 1. С. 38-46.

3. **Карпова Н.С., Суринов А.Е., Ульянов И.С.** Проблемы и возможности использования больших данных в российской статистике // Вопросы статистики. 2016. № 7. С. 3-9.

4. **Плеханов Д.А.** Большие данные и официальная статистика: обзор международной практики внедрения новых источников данных // Вопросы статистики. 2017. № 12. С. 49-60.

5. World Bank, UNSD. Big Data Project Inventory URL: <https://unstats.un.org/bigdata/inventory>.

6. **Hammer C.L., Kostroch D.C., Quirós G.,** STA Internal Group. Big Data: Potential, Challenges, and Statistical Implications. IMF Staff Discussion Note no.17/06. September 2017.

7. A Suggested Framework for the Quality of Big Data. Deliverables of the UNECE Big Data Quality Task Team. December 2014. URL: <http://www1.unece.org/stat/platform/display/bigdata>.

8. **Vale S.** UNECE Task Team on Big Data. Classification of Types of Big Data. 27 June 2013. URL: <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>.

9. Eurostat Big Data Task Force. Item 8 of the agenda. ESS Big Data Action Plan and Roadmap 1.0. Work Programme Objective 11.1. ESSC 2014/22/8/EN - 26/09/2014. 22nd Meeting of the European Statistical System Committee, 26 September 2014, Riga (Latvia).

10. **Моргенштерн О.** О точности экономико-статистических наблюдений. М.: Статистика, 1968. 293 с.

11. **Groves R.M.** A Possible Data Future for the Observational Social Sciences. 2 November 2011. URL:

²⁴ Важность формирования таких «удостоверяющих центров» или «Trusted Data Collaboratives», наряду с развитием международного сотрудничества и созданием глобальной экспериментальной платформы по большим данным, была особо подчеркнута в Декларации по итогам 4-й Международной конференции по большим данным в официальной статистике, состоявшейся 8-10 ноября 2017 года в Боготе (Колумбия). URL: <https://unstats.un.org/unsd/bigdata/conferences/2017/Bogota%20declaration%20-%20Final%20version.pdf>.

²⁵ Дмитрий Медведев. Вступительное слово на заседании Правительственной комиссии по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности 18 декабря 2017 г. URL: <http://government.ru/news/30636>.

<http://studylib.net/doc/11556583/a-possible-data-future-for-the-observational-social-sciences>.

12. **Groves R.M.** Data Production in Organic Data. The Provost's Blog. Weblog. August 19, 2015. URL: <https://blog.provost.georgetown.edu/data-production-in-organic-data>.

13. **Barbosa A.** Innovation and Modernization of National Statistical Systems Through «Trust-

ed Data Collaboratives». Keynote Speech. Proc. of the 4th International UN Conference on Big Data for Official Statistics, 8 November 2017, Bogota, Colombia.

14. **Buono D.** Eurostat Big Data Task Force. Item 4. Smart Statistics. 9 February 2017. URL: https://ec.europa.eu/eurostat/cros/content/item-4-smart-statistics_en.

Информация об авторе

Оксенойт Георгий Константинович - заместитель руководителя Федеральной службы государственной статистики. 107450, г. Москва, ул. Мясницкая, д. 39, стр. 1. E-mail: stat@gks.ru.

References

1. **Oksenoyt G.K.** *The target model - «Statistics-2025»*. Report at the Meeting of the Council on Open Data of the Government Commission for the Coordination of the Open Government, 23 December 2016. (In Russ.) Available from: http://www.gks.ru/free_doc/new_site/rosstat/smi/reliiz-sod.pdf.

2. **Suslov V.I.** Economic Dimensions: Probability and Reliability, Mathematical Modeling, Big Data, E-statistics. *Voprosy statistiki*. 2016;(1):38-46. (In Russ.)

3. **Karpova N.S., Surinov A.Ye., Uliyanov I.S.** Problems and Possibilities for Using Big Data in the Russian Statistics. *Voprosy statistiki*. 2016;(7):3-9. (In Russ.)

4. **Plekhanov D.A.** Big Data and Official Statistics: a Review of International Experience with Integration of New Data Sources. *Voprosy statistiki*. 2017;(12):49-60. (In Russ.)

5. World Bank, UNSD. *Big Data Project Inventory*. Available from: <https://unstats.un.org/bigdata/inventory>.

6. **Hammer C.L., Kostroch D.C., Quirós G.**, STA Internal Group. *Big Data: Potential, Challenges, and Statistical Implications*. IMF Staff Discussion Note no. 17/06. September 2017.

7. A Suggested Framework for the Quality of Big Data. Deliverables of the UNECE Big Data Quality Task Team. December 2014. Available from: <http://www1.unece.org/stat/platform/display/bigdata>.

8. **Vale S.**, UNECE Task Team on Big Data. *Classification of Types of Big Data*. 27 June 2013. Available from: <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>.

9. Eurostat Big Data Task Force. Item 8 of the agenda. ESS Big Data Action Plan and Roadmap 1.0. Work Programme Objective 11.1. ESSC 2014/22/8/EN - 26/09/2014. 22nd Meeting of the European Statistical System Committee, 26 September 2014, Riga (Latvia).

10. **Morgenshtern O.** *On the accuracy of economic observations*. Moscow: Statistika Publ.; 1968. 293 p. (In Russ.)

11. **Groves R.M.** *A Possible Data Future for the Observational Social Sciences*. 2 November 2011. Available from: <http://studylib.net/doc/11556583/a-possible-data-future-for-the-observational-social-sciences>.

12. **Groves R.M.** *Data Production in Organic Data*. The Provost's Blog. Weblog. 19 August 2015. Available from: <https://blog.provost.georgetown.edu/data-production-in-organic-data>.

13. **Barbosa A.** Innovation and Modernization of National Statistical Systems Through «Trusted Data Collaboratives». Keynote Speech. *Proc. of the 4th International UN Conference on Big Data for Official Statistics, 8 November 2017, Bogota, Colombia*.

14. **Buono D.** Eurostat Big Data Task Force. Item 4. Smart Statistics. 9 February 2017. Available from: https://ec.europa.eu/eurostat/cros/content/item-4-smart-statistics_en.

About the author

Oksenoyt G. Konstantinovich - Deputy Head, Federal State Statistics Service. 39, bldg. 1, Miasnitskaya St., Moscow, 107450, Russia. E-mail: stat@gks.ru.