

## Использование информационной инфраструктуры цифровой экономики для повышения качества статистических данных

Официальная статистика является основным поставщиком данных о экономическом состоянии макроэкономического уровня, на основе статистических данных принимается большинство экономических решений государственного масштаба. Работа с данными для органов статистики является ключевым бизнес-процессом. Вместе с тем уровень качества статистических данных, поставляемых Росстатом, не всегда оказывается достаточно высоким. Наблюдаются корректировки статистических данных, выявляются несоответствие между наборами данных, описывающих одно и то же экономическое явление.

**Целью работы** является описание методов сбора и обработки статистической информации, которые будут способствовать повышению качества поставляемых данных.

С информационной точки зрения статистическое агентство занимается организацией информационного обмена между поставщиками данных и потребителями, выступает агрегатором данных. Для организации информационного обмена в рамках сообщества пользователей создается семантическое пространство, призванное обеспечить смысловое наполнение данных. Основную роль в создании семантического пространства играют идентификаторы объектов учета. **В качестве методов** сбора и обработки статистической информации и повышения ее качества в статье рассматриваются использование единых идентификаторов отдельных объектов статистического учета. В международной статистической практике применяются методы стандартизации оборота статистических данных. Информационные стандарты призваны унифицировать идентификаторы и пространство имен для участников оборота статистической информации, что позволяет обеспечить единое семантическое пространство. С применением единых идентификаторов становятся прозрачными процедуры обработки статистических данных, в том числе группировка по разным срезам, а также разложение агрегированных данных на составляющие.

**Результатами** работы являются рекомендации по использованию отдельных элементов информационной инфраструктуры для сбора и анализа статистических данных. В существующей информационной инфраструктуре цифровой экономики существует ряд источников данных, использование которых будет способствовать повышению качества сбора и обработки статистических данных. Для создания семантического пространства статистических данных в РФ наиболее актуальным разделом являются реестры базовых объектов. Использование реестров позволит организовать связывание статистических данных из разных предметных областей, а также реализовать соединение агрегированных данных с микроданными. Существенный прогресс наблюдается в маркировке товаров, которая позволяет отслеживать движение по всем этапам жизненного цикла, а также местоположение объекта. Правительство РФ инициировало проект по маркировке товаров, и эта информация дает возможность получить полное представление о существенной части экономики. Дополнительным информационным источником статистических данных может выступать корпоративный сектор, активно использующий в своей деятельности системы прослеживания, которые выполняют мониторинг товара, транспортных средств, контейнеров, складского хозяйства.

**Заключение:** Существует несколько вариантов обеспечения семантического единообразия статистических данных. Мировой опыт ориентируется на использование веб архитектуры, предполагающей использование технологических идентификаторов. Семантику статистических данных возможно обеспечить путем использования созданного потенциала информационной инфраструктуры, что позволит решить ряд проблем статистического учета.

**Ключевые слова:** информационная модель, производители статистических данных, информационная инфраструктура, идентификатор, маркировка, базовые объекты, реестр

Yuriy P. Lipuntsov

Lomonosov Moscow State University, Moscow, Russia

## Usage of the digital economy information infrastructure to improve the quality of statistical data

Statistics agencies are the main data provider on the economic position of the macroeconomic level. Most economic decisions on a national scale are based on statistical data. Data processing is a key business process for statistical agencies. At the same time, the quality of statistical data supplied by Rosstat is not always high enough. There are adjustments, a discrepancy between data sets describing the same economic phenomenon is revealed. The purpose of the work is to describe the methods of collecting and processing statistical information that will contribute to improving the quality of the presented data.

From the information point of view, the statistical agency is engaged in the organization of information exchange between data providers and consumers, acts as a data aggregator.

To organize the information exchange within community you need to create a semantic space to ensure the meaningful filling of the data. The main role in the semantic space is played by the identifiers of objects. The article considers the unified identifiers of statistical accounting objects as a method of collecting and processing statistical information and improving its quality. The international statistical practice use methods of standardizing the turnover of statistical

data. Information standards are designed to unify identifiers and namespace for participants of the statistical information turnover and to provide a single semantic space. If you use of unified identifiers, the procedures for processing statistical data become transparent, it allow you grouping by different sections, as well as decomposition of aggregated data into components.

**The results** of the work are recommendations on the use of Core component of the information infrastructure for the collection and analysis of statistical data. In the existing information infrastructure of the Russian digital economy, there are a number of data sources, the use of which will improve the quality of collection and processing of statistical data. To create a semantic space of statistical data in the Russian Federation, the most important section is the registers of Core Components. The use of registers will allow you to organize the binding of statistical data from different domains, as well as to implement the link of aggregated data with microdata. Significant progress is observed in the

marking of goods, which allows you to track object's movement through all stages of the life cycle, as well as the location. The government of the Russian Federation initiated a project on labeling of goods, and this information gives an opportunity to get a clear picture of a significant part of the economy. An additional information source of statistical data can be the corporate sector, where actively used tracking systems that monitor the goods, vehicles, containers, warehousing.

**Conclusion:** There are several options for creation of the semantic space for statistical data. World experience is guided by the use of the Web architecture, which involves the technological identifiers. Semantics of statistical data can be ensured by using the potential of the information infrastructure, which will solve a number of problems of statistical accounting.

**Keywords:** Information model, Statistical data producers, Information infrastructure, Identifier, Marking, Core Components, Register

## Введение

Официальная статистическая информация является основой для устойчивого развития как внутри государства, так и для международного сотрудничества. Статистические данные отражают различные сферы в экономической, демографической, социальной и экологической областях. При этом данные статистики должны обладать внутренней согласованностью, сопоставимы по разным регионам, по времени, что позволяет комбинировать данные и одновременно использовать связанные друг с другом данные из разных источников.

Обеспечению согласованности и эффективности статистических систем способствует использование стандартных концепций, классификаций и методов. Для облегчения правильной интерпретации данных статистические ведомства должны предоставлять информацию об источниках, методах и процедурах, а сбор статистических данных осуществляется на основе стандартных определений, единиц, классификаций используемых в различных обследованиях и источниках.

Однако, как указывают эксперты, данные Росстата имеют ряд серьезных изъянов, и их устранение могло бы повысить доверие к публикуемым

ведомством данным [1]. Много вопросов вызывает пересмотр данных. При этом не всегда выполняется одновременное внесение корректировок во все публикации, экспертам не доступна информация о порядке расчета показателей и понятное объяснение причин пересмотра данных.

Качество официальной статистики в значительной мере зависит от граждан, предприятий и других респондентов, которые предоставляют первичные данные. Представители Росстата указывают на низкое качество первичных данных, основные из которых – статистические формы, которые ежемесячно представляют предприятия. Предприятия могут существенно корректировать данные, при этом число корректировок в последнее время растет. В случае, если предприятие не предоставило вовремя отчетность, то сотрудники Росстата самостоятельно заполняют отсутствующие данные, делают собственные оценки, которые могут сильно отличаться от фактических данных.

Помимо этого у пользователей есть вопросы по формату представления данных. Росстат публикует новые данные в форме пресс-релизов, которые содержат основные показатели, но не по всем отраслям. При этом публикуемые файлы не подходят для машинной

обработки. Данные Росстата можно сделать более прозрачными, если разместить в открытом доступе используемые методики. Сейчас во многих случаях экспертам приходится самостоятельно прорабатывать возможные варианты расчета показателей.

Одним из вариантов повышения качества статистических данных министр экономики рассматривает информационные технологии, посредством которых можно добиться взаимосвязки системы статистического учета с другими крупными агрегаторами данных, такими как, например, Банк России и ФНС России. [2]

С развитием информационных технологий статистика становится элементом информационной инфраструктуры государства. С учетом этого официальные статистические данные, имеющие практическую значимость, должны подготавливаться и распространяться статистическими ведомствами на объективной основе, подкрепленной информационными методами и процедурами сбора, обработки, хранения и представления статистических данных [3]. Реализация принципа связанности и согласованности агрегированных данных выполняется на принципах информационного моделирования, отдельные из которых рассмотрим в статье.

## Информационные технологии в работе со статистическими данными

С информационной точки зрения национальные статистические агентства занимаются организацией информационного обмена между участниками организованных сообществ, выступает своеобразным узлом обмена данными (data hub), агрегатором данных. Работа с данными для статистических агентств является ключевым бизнес-процессом и поэтому этот сектор демонстрирует наибольший прогресс в этой области.

С точки зрения пользователя удобно, когда статистические данные из разных предметных областей представлены в сопоставимом, связанном виде. Для этого данные, получаемые из разных источников, должны проходить предварительную обработку на предмет сопоставимости, качества, полноты информации. Выработка методики подготовки связанных данных предполагает тесное сочетание знаний предметной области с применением базовых принципов из области семантических технологий, которые позволяют объединять данные и метаданные, предназначенные для обмена, и хранящиеся в различных форматах с использованием различных сред.

С учетом этого в работе со статистическими данными одним из ключевых моментов деятельности в области данных является стандартизация, «индустриализация производства стандартной статистической информации» [4].

Один из основных принципов, обеспечивающий качество данных — это согласованность, которая подразумевает, что данные описывают один и тот же объект, и поступающие из разных источников легко, а в идеале — автоматически, сопоставляются. Здесь возникает два направления, по которым необходимо организовать со-

поставимость: первое — сопоставить объекты, а второе — сопоставить характеристики, описывающие объект.

Сопоставимость статистических данных — это трудная задача, поскольку определенная часть данных статистических отчетов готовится вручную и не предполагает интерпретацию первичных данных. С появлением специальных статистических форматов данных, которые применяются при составлении статистической отчетности возрастает повторное использование данных. Однако в связи со сложностью использования этих форматов конечными пользователями возникает барьер, поскольку основная часть этих пользователей ориентирована на такие форматы, как Microsoft Excel или CSV. У пользователей есть навыки по работе с данными в этих форматах, для обработки таких данных можно найти инструменты и библиотеки на любом языке программирования. Обратная сторона таких широко распространенных форматов — ограниченность использования аннотаций и метаданных, которые позволяют правильно интерпретировать данные. Возникают сложности с обновлением таких данных. Более того, при использовании таких форматов данные и метаданные обычно отделены друг от друга. Результаты исследований показывают, что при работе с этими форматами 80% времени тратится на приведение данных из разных источников к пригодному для обработки виду и только 20% на саму обработку данных и выводы. Поэтому перспективным вариантом развития статистической информатики представляются форматы связанных данных.

Один из значимых проектов по разработке совокупности стандартов по обмену статистическими данными и метаданными является инициатива SDMX (Statistical Data and

Metadata eXchange) [5], которая призвана содействовать обмену статистическими данными и метаданными с использованием современных информационных технологий. Этот формат в 2005 году принят в качестве международного стандарта ISO: TS 17369 (SDMX) [6].

SDMX это логическая модель для описания статистических данных, которая содержит рекомендации по структурированию содержания и позволяет реализовать стандарт для автоматического взаимодействия между информационными системами.

Информационная модель SDMX включает широкий набор формальных объектов, которые описывают участников, процессы и ресурсы, участвующие в статистическом обмене и включает в себя два раздела: данные и метаданные.

При использовании SDMX формата, то есть в формате связанных данных, можно сохранять метаданные вместе с данными, что позволяет обеспечить машиночитабельность как данных, так и метаданных. Поэтому одна из задач модели выступать в качестве механизма гарантии точности связи между данными и метаданными, которые могут быть представлены в различных форматах.

Процесс статистического обмена моделируется шаг за шагом, фиксируется вся информация, в том числе о поставщиках данных и справочных метаданных. Это позволяет полноценно управлять информационным обменом.

Работоспособность технических стандартов, в сочетании с принципами, которые связаны с содержанием и семантикой обмена статистической информацией в рамках SDMX обеспечивается путем гармонизации контента. Для качественного информационного обмена между ведомствами необходимо, чтобы все участники обмена использовали единое описание как базовых информационных объектов,

так и информационных объектов предметных областей. В рамках SDMX ставится задача описания данных, посредством детального определения всех элементов описания, такие как списки кодов, концептуальные схемы и концепции, наборы данных и т.д. Посредством такого определения для пользователей SDMX системы создаются условия для однозначного толкования данных. С точки зрения архитектуры информационной системы идеология SDMX предполагает создание централизованного репозитория, в рамках которого определено пространство имен.

Расширением идеологии SDMX является использование веб-архитектуры для обнаружения наборов данных. В веб-архитектуре в качестве идентификаторов используется HTTP URI. В среде веб-статистическая информация может быть найдена путем назначения идентификаторов URI для всех экземпляров модели данных SDMX. На этих принципах создан Статистический словарь – The Statistical Core Vocabulary (SCOVO) (SCOVO) [7]. Это RDF словарь для представления статистических данных, который позволяет организовать обмен данными между производителями и потребителями. Помимо этого, его можно комбинировать с другими словарями RDF.

Это словарь изначально разработан в рамках Проекта RISEE ( RDFizing and Interlinking the EuroStat Data Set Effort) [8]. Он также использовался для представления статистических данных в формате наборов данных RDF в проектах правительства Великобритании [9].

#### **Возможности по повышению качества статистических данных**

Основная информационная задача, которую нужно решить в рамках организации

сбора статистических данных и представления ее пользователям можно определить как предоставление семантического пространства для участников информационного обмена. Семантика в информационном пространстве определяется двумя составляющими:

- идентификация объектов;
- пространство имен.

Идентификация объектов среди этих двух компонент является более важной составной частью.

В разделе обзора литературы было представлено как решается задача определения пространства имен в модели SDMX, определяющей списки кодов, концепции и прочие элементы семантики. В модели SCOVO проблемы семантики решались путем назначения глобальных идентификаторов всем объектам модели обмена статистическими данными: поставщикам данных, индикаторам, измерителям и т.д. Веб это международное информационное пространство и построение универсальной модели, используемой для всех случаев поставки и потребления статистических данных задача интересная, но не такая простая. Одно дело принять международные стандарты, которые определяют рамочную модель для организации информационного обмена, другое дело реализация такой модели в виде логической и физической модели, где должны быть учтены все нюансы учетной политики статистических агентств отдельной страны с ее множеством особенностей.

Поэтому попробуем сформулировать принципы, на которых может быть основан сбор статистических данных с учетом имеющихся в информационной инфраструктуре элементов. Основу семантики задают идентификаторы объектов, задействованных в информационном обмене. Идентификаторы объектов создаются в регистрах, разра-

ботанных, как правило, для отдельных объектов. Обычно это государственные системы, в которых отслеживается актуальное состояние отдельных объектов: физических лиц, юридических лиц, объектов недвижимости и т.д. Все операции протекающие с этими объектами, должны получать информацию об объекте из этого источника. Регистр предоставляет актуальную информацию об объекте, в том числе его идентификатор.

Пространство имен (Name Spaces) в случае статистического обмена позволяет участникам информационного обмена согласовывать правила описания отдельных элементов. С использованием пространства имен участники информационного обмена могут получать информацию от других участников, выполнять преобразования этих данных, а также публиковать данные с результатами своей деятельности для внешних потребителей данных.

Таким образом пространство имен предоставляет возможность описать данные на понятном для всех участников информационного пространства формате.

#### **3.1. Роль идентификаторов в информационном обмене**

Создание информационной инфраструктуры обмена данными, включающей большое количество организаций и систем, предполагает однозначное понимание передаваемой информации всеми участниками обмена. Ключевым элементом в этом случае выступают идентификаторы. Эмитентами идентификаторов, используемых для разных категорий объектов, может выступать один центр или совокупность центров кодификации. Идентификаторы, назначенные в одном центре, могут встречаться и повторно использоваться в другом месте, если не предусмотреть систему консультаций между эмитентами

идентификаторов. Необходима гарантия того, что результаты деятельности одного эмитента кодов будут известны другим участникам. Взаимодействие между многими системами требует разработки идентификаторов, позволяющих использовать их в сервисах за пределами прямого управления [10].

Обеспечение надежной и устойчивой инфраструктуры информационного обмена предполагает, что идентификатор обладает следующими характеристиками:

- Устойчивость: идентификатор должен быть неизменным и постоянным и не подлежит изменению ни при каких обстоятельствах.

- Расширяемость: идентификаторы, которые будут использоваться в информационном обороте потребители и производители данных, могут быть выпущены разными органами, поддерживаться различными декларациями метаданных. Все идентификаторы в рамках единой информационной среды объединяются на основе согласованной модели данных [11].

Методология использования идентификаторов отдельных предметных областей для организации межведомственного информационного оборота на примере финансовой сферы рассмотрены в работе [12].

Наряду с независимыми от языка, уникальными, постоянными и расширяемыми идентификаторами для организации статистического обмена нужны дополнительные данные, которые могут быть использованы для обозначения концепций и предоставления информации о концепциях, т.е. метаданные, стандартизированные в SDMX.

### **3.2. Идентификаторы как часть информационной инфраструктуры цифровой экономики**

Производителями первичных данных, поставщиками данных для органов статистики

являются граждане, предприятия, государственные органы и прочие организации. Для организации сбора информации критичным является наличие полного перечня объектов статистического учета, будь то полный список предприятий отрасли, транспортных средств, товаров и прочих объектов. В реальной экономике происходит постоянные изменения, например с объектами «Юридическое лицо» происходят слияния, поглощения, ликвидация, ребрендинг и прочие операции, смысл которых должен проинтерпретировать специалист предметной области. Это не входит в задачи статистического учета, но от правильной интерпретации зависит качество статистической информации. Поставщиками актуальной информации в условиях создающейся инфраструктуры цифровой экономики являются реестры, которые в данном случае выступают агрегатором актуальных данных относительно отдельного объекта учета.

Федеральная налоговая служба (ФНС) ведет реестр юридических лиц, реестр объектов недвижимости ведется Федеральной кадастровой палатой. ФНС в скором времени запустит единый государственный реестр актов гражданского состояния, который станет основой для создания единого Реестра населения.

Критическими для поставки статистических данных элементами информационной инфраструктуры, являются регистры не только универсальных объектов, таких как юридические и физические лица, объекты недвижимости, а также регистры предметных областей – транспортные средства, объекты образовательной, медицинской деятельности, объекты железнодорожной, энергетической, коммуникационной инфраструктуры и т.д. [13]

Статистические данные как правило имеют территориаль-

ную привязку, которая должна иметь единый формат описания местоположения объектов, одним из вариантов которой может выступать единая адресная система ФИАС, а в перспективе можно рассматривать вариант с использованием данных навигационной системы, которая используется для осуществления хозяйственных транзакций. Это позволит перейти от позиционирования крупных компаний по месту юридического адреса к местам дислокации добычи, производства, переработки, распределения, что также скажется на качестве статистических данных. Включение навигационной спутниковой системы «ГЛОНАСС» в информационную инфраструктуру цифровой экономики для обеспечения единого описания местоположения в пространстве, а также использования единого времени, формируемое по сигналам ГНСС, способно повысить качество информационного обеспечения в разных секторах экономики.

Использование единых идентификаторов, стандартного формата времени и места позволят повысить сопоставимость данных поступающих из разных источников, что скажется на качестве агрегированных, в том числе статистических данных.

Помимо регистров, осуществляющих учет объектов, дополнительным источником первичных данных из информационной инфраструктуры цифровой экономики можно рассматривать операторов цифровой маркировки товаров. Использование бар-кодов при розничной покупке товаров в магазине является привычным явлением. Посредством такой маркировки розничные продавцы автоматизируют процесс определения цены и прочих атрибутов товаров. Расширение идеологии маркировки на отдельные группы товаров, присваивание им уникальных

кодов позволит решить многие задачи регулирования оборота отдельных групп товаров, в том числе и повысить качество статистической информации по этим группам.

Правительство России утвердило перечень из десяти групп товаров, подлежащих обязательной маркировке<sup>1</sup>. В список попали табачная продукция, шины и покрышки, обувные товары, белье и прочие товары. Сейчас в России маркируются меховые изделия и действует пилотный проект маркировки лекарств, драгоценных металлов и камней.

На настоящий момент маркировкой отдельных товарных групп занимаются разные ведомства: шуб — Федеральная налоговая служба, лекарств — Росздравнадзор. До 2024 года все вопросы маркировки по широкому спектру номенклатуры должны перейти к единому оператору. В настоящий момент утвержден единый оператор маркировки, который «готов включиться в работу с любыми товарными группами» [14], и взаимодействует с молочным рынком, производителями пива и минеральной воды, парфюмерии и шин.

Система маркировки позволяет отслеживать состояние объекта в цепочках поставок и всех событиях, происходящих на отдельных этапах жизненного цикла объекта, в том числе:

- происхождению материалов и деталей;
- истории обработки;
- распределению и расположению продукта или услуги после доставки.

В корпоративном секторе для решения вопросов управления деятельностью нашла широкое применение идеология Прослеживаемость. Под прослеживаемостью понимают возможность отслеживать

историю, приложение или местоположение объекта [15]. При рассмотрении продукта или услуги отслеживание может относиться к:

- Переработка сырья, ингредиентов, получение промежуточных продуктов, компонентов и изготовление конечного продукта;

- Агрегирование и дезагрегация продуктов и связь с активами (например, возвращаемые активы).

- Транспортировка и распределение, включая трансграничную торговлю.

- Обслуживание, ремонт и капитальный ремонт в течение нескольких циклов использования продукта.

- Потребление продуктов, включая выдачу.

- Уничтожение продукта и утилизация материалов.

Прослеживаемый объект представляет собой физический или цифровой объект, для которого необходимо получить информацию об истории, использовании или местоположении. В рамках одного процесса отслеживаемый объект может проходить трансформация объекта, один объект может преобразовываться в другой. Система отслеживания ориентирована на то, чтобы учитывать все особенности производственных и других особенностей предметной области и корректно отражать все возможные нюансы.

Примеры прослеживаемых объектов логистики включают в себя продукты (например, товары народного потребления, лекарства, электронные устройства), логистические единицы (например, товары на полетах, посылки) и активы (например, грузовые автомобили, суда, поезда, вилочные подъемники).

Реализацией идеологии прослеживаемости на практике занимается Ассоциация автоматической идентификации GS1<sup>2</sup>, независимая не-

коммерческая организация по созданию и внедрению стандартов.

Органам статистики возможно следует организовать взаимодействие с ассоциацией и получать от них данные для поведения оценок, которые могут быть хорошим дополнением для регламентной отчетности, предоставляемой предприятиями на регулярной основе.

### 3.3. Использование идентификаторов в отдельных статистических предметных областях

Разделы информационной инфраструктуры цифровой экономики, которые можно использовать при сборе статистических данных представлены в таблице 1.

В разделе «Статистические предметные области» модели SDMX представлен список предметных областей статистической информации, используемый для организации статистического обмена и категоризации. Этот список можно использовать в качестве схемы для описания категорий для различных контрагентов, поставляющих данные и метаданные. Отдельными поставщиками данных могут выступать разделы информационной инфраструктуры, представленные в таблице 1. Используя список предметных областей организаторы отдельного направления могут включать в стандартные структуры данных, определения, понятия элементы инфраструктуры.

Список предметных областей разделен на три направления: демография и социальная статистика, экономика и окружающая среда и межотраслевая статистика. Дополним этот список разделами информационной инфраструктуры, который может быть использован для расчета показателей.

<sup>1</sup> Распоряжение Правительства РФ от 28 апреля 2018 г. № 792-п <http://static.government.ru/media/files/Srq8pmDMRTAVFxN8aWnpOmCVxGm9gZil.pdf>

<sup>2</sup> [www.gs1.ru.org/](http://www.gs1.ru.org/)

Таблица 1

## Разделы информационной инфраструктуры и их содержание

Раздел инфраструктуры	Содержание
Универсальные реестры (УР)	Базовые объекты, представленные во всех сферах деятельности: юридические лица, физические лица, объекты недвижимости, местоположение
Реестры предметных областей (РПО)	Базовые объекты отдельных сфер деятельности, пример: транспортные средства, объекты железнодорожной, энергетической, коммуникационной инфраструктуры, объекты образовательной, медицинской деятельности и т.д.
Маркировка товаров (МТ)	Табачная продукция, духи и туалетная вода, шины и покрышки, обувные товары, белье, фотокамеры, фотовспышки, отдельные предметы одежды, меховые изделия, лекарства, драгоценные металлы и камни
Автоматическая идентификация (АИ)	Категории стандарта GS1: Продукты (например, товары народного потребления, лекарства, электронные устройства), логистические единицы (например, товары на полетах, посылки) и активы (например, грузовые автомобили, суда, поезда, вилочные подъемники)

Использование идеологии идентификаторов будет способствовать повышению качества статистических данных как на этапе сбора, так и на этапе их анализа.

## 3.4. Связь агрегированных данных с микро-данными

В работе [16] приведено деление пользователей статистики на три категории: туристы, фермеры и шахтеры.

Туристы являются широкой публикой и заинтересованы в агрегированной информации по отдельной предметной области.

Фермеры являются экспертами предметной области, которые интересуются широким спектром статистических данных, как правило, на уровне совокупности. Их интересует временной срез, региональный.

Шахтеры являются статистиками, исследователями и аналитиками, которые глубоко погружаются в определенную предметную область на уровне микроданных.

У каждого другого пользователя есть предпочтения в отношении того, как представлены статистические данные и фактические данные.

Туристам нравится, таблицы, графики, истории и визу-

Таблица 2

## Список предметных областей и разделы информационной инфраструктуры для расчета индикаторов

Демографическая и социальная статистика	Раздел ИИ	Экономическая статистика	Раздел ИИ	Окружающая среда и Многоотраслевая статистика	Раздел ИИ
1.1. Население и миграция	УР	2.1. Макроэкономическая статистика	УР, РПО, МТ, АИ	3.1. Окружающая среда	УР, РПО, АИ
1.2. Труд	УР	2.2. Экономические счета	УР, РПО, МТ, АИ	3.2. Региональная статистика и статистика малых районов	УР, РПО, АИ
1.3. Образование	РПО	2.3. Бизнес-статистика	УР, РПО, МТ, АИ	3.3 Многоотраслевая статистика и показатели	УР, РПО, МТ, АИ
1.4. Здоровье	РПО	2.4. Отраслевая статистика	УР, РПО, МТ, АИ	3.3.1. Условия жизни, бедность и прочие социальные вопросы	УР, РПО, АИ
1.5. Доходы и потребление	УР, РПО	2.4.1. Сельское хозяйство, лесное хозяйство, рыболовство	УР, РПО, МТ, АИ	3.3.2. Специфичные группы населения	УР, РПО
1.6. Социальная защита	РПО	2.4.2. Энергия	УР, РПО, МТ, АИ	3.3.3. Информационное общество	УР, РПО, МТ, АИ
1.7. Жилье	УР	2.4.3. Добыча, производство, строительство	УР, РПО, МТ, АИ	3.3.4. Глобализация	УР, РПО, МТ, АИ
1.8. Правосудие и преступность	РПО	2.4.4. Транспорт	УР, РПО, АИ	3.3.5. Индикаторы, относящиеся к Целям развития тысячелетия	
1.9. Культура	РПО	2.4.5. Туризм	УР, РПО, АИ	3.3.6. Устойчивое развитие	УР, РПО
1.10. Политические и прочие сообщества	РПО	2.4.6. Банковская, страховая, финансовая статистика	УР, РПО	3.3.7. Предпринимательство	УР, РПО, МТ, АИ
1.11. Времяпровождение	РПО	2.5. Государственные финансы, фискальная статистика и статистика государственного сектора	УР, РПО	3.4. Ежегодники и другие сборники	УР, РПО, МТ, АИ
		2.6. Международная торговля и платежный баланс	УР, РПО, МТ, АИ		
		2.7. Цены	УР, РПО, МТ, АИ		
		2.8. Затраты на оплату труда	УР, РПО, МТ, АИ		
		2.9. Наука, технологии и инновации	УР, РПО		

ализации на бумаге и в Интернете, которые помогают им понять данные, такие как привлекательные презентации демографических тенденций.

Фермеры предпочитают наборы данных, совокупность совокупных данных с течением времени, которые хорошо классифицированы, отформатированы и доступны для компьютерных процессов обработки данных, например в форме базы данных.

Шахтеры предпочитают иметь дело с первичными, необработанными данными и просматривать данные в формате, который позволяет выполнять разнообразный анализ, например анонимные микроданные переписи.

Наиболее сложный с информационной точки зрения вопрос с последней категорией пользователей, которые для проведения своих исследований претендуют на доступ к микроданным.

Преобразующую роль в организации работы с данными на таком уровне играют идентификаторы. В частности, идентификаторы поставщиков данных, которые позволят просматривать микроданные. В этом случае вопрос с корректировками на агрегированном уровне может быть рассмотрен на уровне отдельных поставщиков данных и снимется вопрос о причинах таких корректировок.

Помимо использования идентификаторов необходимо документировать метаданные согласно стандартизованным системам метаданных, пользователи должны получать информацию по методологии статистических процессов, в том числе об использовании административных данных.

В настоящее время Росстат не может использовать данные ФНС России в той же степени, как это осуществляется в других странах ОЭСР. «Это связано с отказом ФНС России предоставлять пообъектные данные о

результатах деятельности налогоплательщиков, что не позволяет формировать статистические показатели в соответствии с требованиями ОЭСР.

Ряд министерств и ведомств, располагающих информационными ресурсами, необходимыми для построения национальных счетов, а также баланса активов и пассивов отказываются от ее представления в Росстат, ссылаясь на обеспечение конфиденциальности или отсутствие поручений Правительства Российской Федерации.» [17]

В целях повышения качества официальной статистической информации, снижения нагрузки на респондентов, экономии бюджетных средств требуется радикальное расширение доступа органов государственной статистики к пообъектным и сводным данным, имеющимся, прежде всего, в ФНС России, Пенсионном фонде Российской Федерации, Фонде социального страхования Российской Федерации.

### 3.5. Предложения по отчётности

Важным элементом предложенной системы оценки эффективности является наполненность и представленность информационной базы. С учетом этого целесообразно рассмотреть инициативу по созданию форм статистической отчетности, которая бы включала отчет субъектов хозяйственной деятельности об использовании государственной информационной инфраструктуры для организации информационного оборота как в регламентированной деятельности, такой как предоставление отчета в регулирующие органы государственной власти, так и использование информационных ресурсов и стандартов для организации взаимодействия в процессе хозяйственной деятельности: информационный обмен с поставщиками, подрядчиками,

клиентами, кадровыми агентствами, и прочими участниками. Список предметных областей модели SDMX, приведенных в таблице 2 содержит раздел 3.3.3 Информационное общество, в который могут быть включены эти формы отчетности.

С одной стороны это будет повышать качество статистических данных, с другой – способствовать росту информационной зрелости предприятий.

### Заключение

В статье показаны тенденции в области информационного моделирования, предполагающие активное использование внешних данных. Современное предприятие независимо от сферы деятельности предполагает активное информационное взаимодействие системных компонент, их интеграцию. Для управления такой организацией необходимы данные из различных предметных областей, ее информационные системы ориентированы на взаимодействие с другими системами, а информационная модель включает несколько источников данных и публикует данные в формате понятном для окружающих пользователей.

С технологической точки зрения наиболее перспективная форма информационного обмена между многими участниками – стандартизация состава данных, подлежащих обмену. Для создания информационной инфраструктуры цифровой экономики многообразий сфер и направлений, для которых необходимо стандартизировать информационный оборот можно классифицировать по направлениям деятельности организации и разработать шаблоны данных для этих направлений.

С управленческой точки зрения информационная инфраструктура предоставляет возможность получить каче-

ственный информационный обмен между участниками различных сфер экономической деятельности. Получение качественной информации-

ной поддержки способно существенно повысить качество управленческих решений как на оперативном, так и на стратегическом уровне. В статье

показаны основные элементы информационной инфраструктуры, которые необходимы для функционирования цифровой экономики.

## Литература

1. Салихов М. В заложниках у статистической погрешности: почему сложно верить Росстату 27.06.2018. URL: <https://www.rbc.ru/opinions/economics/27/06/2018/5b3201229a794725025a6958>. (Дата обращения: 07.07.2018).
2. ТАСС. «МЭР: Росстат должен повысить качество статистики и уровень доверия к ней» 04.04.2017. URL: <http://tass.ru/ekonomika/4153390>. (Дата обращения: 18.05.2018).
3. UNSD. «Official Statistics: Principles and Practices» 18.02.2017. URL: <https://unstats.un.org/unsd/methods/statorg/FP-Russian.pdf>. (Дата обращения: 19.04.2018).
4. M. Pellegrino. «Maintaining the quality of EU statistics while enabling re-use» в SEMIC, Dublin, 2013.
5. Statistical Working Group. «Statistical Data and Metadata Exchange» URL: <http://sdmx.org/>. (Дата обращения: 23 3 2015).
6. ISO/TC 154 , «ISO/TS 17369:2005 Statistical data and metadata exchange (SDMX)» 10 10 2005. URL: <https://www.iso.org/standard/40555.html>. (Дата обращения: 16 05 2018).
7. M. Hausenblas, W. Halb, Y. Raimond, L. Feigenbaum, D. Ayers « SCOVO: Using Statistics on the Web of Data» в Proceedings of ESWC 2009 — 6th European Semantic Web Conference, Heraklion, 2009.
8. M. Hausenblas, W. Halb и Y. Raimond, «Scripting User Contributed Interlinking» в 4th Workshop on Scripting for the Semantic Web (SFSW08). Tenerife, 2008.
9. K. Alexander, R. Cyganiak, M. Hausenblas и J. Zhao, «Describing Linked Datasets» в Proceedings of the Linked Data on the Web Workshop (LDOW2009), Madrid, 2009.
10. W. Arms, Digital Libraries. Retrieved 04.04.2017. Boston: M.I.T. Press., 2000 .
11. N. Paskin Naming and meaning of Digital Objects // Proceedings of the 2nd International Conference on Automated Production of Cross Media, 2006.
12. Y. Lipuntsov, R. Beatch, I. Collier. Financial Markets Data Collection Using the Information Model of Interagency Cooperation and the International System of Codification of Financial Instruments // Communications in Computer and Information Science. 2017. т. 745.
13. Y. Lipuntsov. An Information model of Interagency Communication Based on Distributed Data Storage» в International Conference on

## References

1. Salikhov M. V zalozhnikakh u statisticheskoy pogreshnosti: pochemu slozhno verit' Rosstatu 27.06.2018. URL: <https://www.rbc.ru/opinions/economics/27/06/2018/5b3201229a794725025a6958>. (Accessed:07.07.2018). (In Russ.)
2. TASS. «MER: Rosstat dolzhen povysit' kachestvo statistiki i uroven' doveriya k ney» 04.04.2017. URL: <http://tass.ru/ekonomika/4153390>. (Accessed:18.05.2018). (In Russ.)
3. UNSD. «Official Statistics: Principles and Practices» 18.02.2017. URL: <https://unstats.un.org/unsd/methods/statorg/FP-Russian.pdf>. (Accessed: 19.04.2018).
4. M. Pellegrino. Maintaining the quality of EU statistics while enabling re-use. SEMIC, Dublin, 2013.
5. Statistical Working Group. «Statistical Data and Metadata Exchange» URL: <http://sdmx.org/>. (Accessed: 23.03.2015).
6. ISO/TC 154 , «ISO/TS 17369:2005 Statistical data and metadata exchange (SDMX)» 10 10 2005. URL: <https://www.iso.org/standard/40555.html>. (Accessed: 16 05 2018).
7. M. Hausenblas, W. Halb, Y. Raimond, L. Feigenbaum, D. Ayers « SCOVO: Using Statistics on the Web of Data». Proceedings of ESWC 2009 — 6th European Semantic Web Conference, Heraklion, 2009.
8. M. Hausenblas, W. Halb и Y. Raimond, «Scripting User Contributed Interlinking». 4th Workshop on Scripting for the Semantic Web (SFSW08). Tenerife, 2008.
9. K. Alexander, R. Cyganiak, M. Hausenblas и J. Zhao, «Describing Linked Datasets» в Proceedings of the Linked Data on the Web Workshop (LDOW2009), Madrid, 2009.
10. W. Arms, Digital Libraries. Retrieved 04.04.2017. Boston: M.I.T. Press., 2000 .
11. N. Paskin Naming and meaning of Digital Objects // Proceedings of the 2nd International Conference on Automated Production of Cross Media, 2006.
12. Y. Lipuntsov, R. Beatch, I. Collier. Financial Markets Data Collection Using the Information Model of Interagency Cooperation and the International System of Codification of Financial Instruments // Communications in Computer and Information Science. 2017. Vol. 745.
13. Y. Lipuntsov. An Information model of Interagency Communication Based on Distributed Data Storage» в International Conference on Elec-

Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '16), New York, 2016.

14. РБК. Назван единый оператор по электронной маркировке товаров» 12.07.2018. URL: <https://www.rbc.ru/rbcfreenews/5b45ee149a79474230d4c946?from=newsfeed>. (Дата обращения: 08.08.2018).

15. GS1. «GS1 General Specifications» URL: [https://www.gs1.org/docs/barcodes/GS1\\_General\\_Specifications.pdf](https://www.gs1.org/docs/barcodes/GS1_General_Specifications.pdf).

16. R. Cyganiak, M. Hausenblas, E. McCuire. Official Statistics and the Practice of Data Fidelity. New York: Linking Government Data, 2011.

17. Росстат. «О некоторых дополнительных мерах по реализации государственной политики в сфере государственной статистической деятельности в связи с вступлением Российской Федерации в ОЭСР» 2017. URL: [www.gks.ru/free\\_doc/new\\_site/rosstat/os/doc12\\_5.doc](http://www.gks.ru/free_doc/new_site/rosstat/os/doc12_5.doc). (Дата обращения: 03.05.2018).

18. UK Cabinet Office «e-Government Metadata Standard» 28.08.2006. URL: [www.nationalarchives.gov.uk/documents/information-management/egms-metadata-standard.pdf](http://www.nationalarchives.gov.uk/documents/information-management/egms-metadata-standard.pdf).

19. EuroStat. Кодекс норм европейской статистики URL: <http://ec.europa.eu/eurostat/quality>.

20. R. Cyganiak, S. Field, A. Gregory и J. Tennison. Semantic Statistics: Bringing Together SDMX and SCOVO» в Richard Cyganiak Simon Field Simon Field Arofan Gregory Jeni Tennison, Semantic SProceedings of the WWW2010 Workshop on Linked Data on the Web, LDOW 2010, Raleigh, 2010.

Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '16), New York, 2016.

14. RBK. Nazvan edinyy operator po elektronnoy markirovke tovarov 12.07.2018. URL: <https://www.rbc.ru/rbcfreenews/5b45ee149a79474230d4c946?from=newsfeed>. (Accessed: 08.08.2018). (In Russ.)

15. GS1. «GS1 General Specifications» URL: [https://www.gs1.org/docs/barcodes/GS1\\_General\\_Specifications.pdf](https://www.gs1.org/docs/barcodes/GS1_General_Specifications.pdf).

16. R. Cyganiak, M. Hausenblas, E. McCuire. Official Statistics and the Practice of Data Fidelity. New York: Linking Government Data, 2011.

17. Rosstat. «O nekotorykh dopolnitel'nykh merakh po realizatsii gosudarstvennoy politiki v sfere gosudarstvennoy statisticheskoy deyatel'nosti v svyazi s vstupleniyem Rossiyskoy Federatsii v OESR» 2017. URL: [www.gks.ru/free\\_doc/new\\_site/rosstat/os/doc12\\_5.doc](http://www.gks.ru/free_doc/new_site/rosstat/os/doc12_5.doc). (Accessed: 03.05.2018). (In Russ.)

18. UK Cabinet Office «e-Government Metadata Standard» 28.08.2006. URL: [www.nationalarchives.gov.uk/documents/information-management/egms-metadata-standard.pdf](http://www.nationalarchives.gov.uk/documents/information-management/egms-metadata-standard.pdf).

19. EuroStat. Kodeks norm evropeyskoy statistiki URL: <http://ec.europa.eu/eurostat/quality>.

20. R. Cyganiak, S. Field, A. Gregory и J. Tennison. Semantic Statistics: Bringing Together SDMX and SCOVO» в Richard Cyganiak Simon Field Simon Field Arofan Gregory Jeni Tennison, Semantic SProceedings of the WWW2010 Workshop on Linked Data on the Web, LDOW 2010, Raleigh, 2010. (In Russ.)

#### Сведения об авторе

**Юрий Павлович Липунцов**

К.э.н., доцент, доцент Кафедры экономической информатики

Московский государственный университет

имени М.В. Ломоносова,

Москва, Россия

Эл. почта: [lipuntsov@econ.msu.ru](mailto:lipuntsov@econ.msu.ru)

#### Information about the author

**Yuriy P. Lipuntsov**

Cand. Sci. (Economics), Associate Professor, Associate Professor of the Department of Economic Informatics

Lomonosov Moscow State University,

Moscow, Russia

E-mail: [lipuntsov@econ.msu.ru](mailto:lipuntsov@econ.msu.ru)