

КАКОЙ ДОЛЖНА БЫТЬ РОССИЙСКАЯ ИНФОРМАЦИОННО-СТАТИСТИЧЕСКАЯ СИСТЕМА?*

В.А. Бессонов

Информационно-коммуникационные технологии достигли такого уровня развития, что официальные сайты статистических служб и доступные на них информационные ресурсы стали основным каналом распространения статистических данных. Поэтому обсуждать проблемы распространения статистической информации в настоящее время следует в контексте баз данных с доступом через Интернет, то есть информационно-статистических систем. Переход от статистических сборников к таким системам способен привести к кардинальному увеличению объема и повышению качества доступной пользователям статистической информации одновременно с резким снижением издержек доступа к данным.

На протяжении ряда лет пользователям доступны официальный сайт Росстата, Центральная база статистических данных и Единая межведомственная информационно-статистическая система, в совокупности претендующие на роль российской информационно-статистической системы. Вместе с тем имеются основания полагать, что указанные информационные ресурсы устраивают пользователей далеко не в полной мере.

Цель настоящей работы - сформулировать соображения автора о том, какая информационно-статистическая система нужна пользователям, решающим задачи анализа российской экономической динамики.

В статье обсуждаются требования к такой системе и соображения по организации процесса ее создания, сопровождения и развития. Рассматриваются информационные структуры, на которые должна быть ориентирована система, связи между ними, отражение общей логики всей совокупности статистических данных в базе данных. Делается упор на обеспечение требований полноты, избыточности, непротиворечивости, актуальности, точности. Обсуждается состав нечисловой информации в базе данных. Рассматриваются требования к функциональным возможностям системы, вопросы ее создания, сопровождения и развития. В силу того что создание системы имеет значительную экономико-статистическую составляющую, она не может быть создана лишь силами специалистов по информационным технологиям. Обосновывается необходимость организации взаимодействия разработчиков системы с ее потенциальными пользователями.

Ключевые слова: база данных, временной ряд, доступ к данным, экономическая статистика, экономический индекс.

JEL: C82, C88.

На протяжении многих десятилетий функционирования национальных статистических служб основным каналом распространения их результатов являлись статистические сборники и другие печатные издания. С развитием информационно-коммуникационных технологий (ИКТ), и в первую очередь Интернета, все более важное значение стали приобретать официальные сайты статистических служб. Наконец, ИКТ достигли такого уровня развития, а доступ к Интернету стал настолько простым, удобным, дешевым, быстрым и повсеместным, что именно официальные сайты и доступные на них информационные ресурсы стали основным каналом получения статистических данных для подавляющего большинства пользователей.

Роль статистических сборников и других печатных изданий в распространении информации изменилась и стала второстепенной. В передовых странах из основного источника данных они трансформировались в полезные вспомогательные материалы, содержащие наиболее актуальную статистическую информацию в виде обширного набора тщательно продуманных аналитических таблиц, служащих своего рода полуфабрикатом, подспорьем для анализа ситуации в экономике. Да и эти материалы становятся доступными большинству пользователей не в бумажном виде, а в электронном - посредством официальных сайтов статистических служб.

Соответственно, обсуждать проблемы распространения статистической информа-

Бессонов Владимир Аркадьевич (bessonov@hse.ru) - канд. физ.-мат. наук, заведующий Лабораторией исследования проблем инфляции и экономического роста Экспертного института НИУ ВШЭ (г. Москва, Россия).

** Статья написана в рамках проекта, финансируемого Программой фундаментальных исследований НИУ ВШЭ. Автор выражает признательность Э.Ф. Баранову за полезные замечания.*

ции в настоящее время следует в контексте баз данных с доступом через Интернет, то есть информационно-статистических систем (ИСС). Такие информационные ресурсы становятся не только основным средством доступа к статистическим данным, но они существенно расширяют и возможности аккумулирования информации за длительные интервалы времени. В развитых странах для этого ранее публиковались специальные издания по исторической статистике¹. Современные информационно-статистические системы предоставляют в этом плане несопоставимо более развитые возможности, поскольку они позволяют организовать накопление и доступ к данным любой структуры и любого объема, в отличие от печатных изданий, имеющих присущие им ограничения по размеру листа бумаги и по числу страниц.

Помимо этого, в отличие от дорогостоящих переизданий раз в несколько лет сборников исторической статистики, информационно-статистические системы допускают проведение актуализации содержащихся в них данных в режиме реального времени, по мере поступления новых данных и уточнения ретроспективных оценок, без привязки к срокам выхода твердой копии, постоянно обеспечивая пользователям доступ к полным и актуальным данным на максимально возможную глубину по времени. В силу отсутствия ограничений на объем хранимой информации такие системы способны обеспечивать доступ не только к данным годовых итогов, но и к показателям на квартальной и месячной основе. Таким образом, они позволяют стереть грань между источниками оперативных данных и данных исторической статистики, обеспечивая постоянный доступ ко всем накопленным массивам данных.

Переход от печатных изданий к полноценным ИСС приводит к кардинальному увеличению объема и повышению качества доступной пользователям статистической информации одновременно с резким снижением издержек доступа к данным. Следствием такого перевода интерфейса официальной статистики на принципиально иной уровень является повышение качества исследований и аналитики, что способствует

лучшему пониманию происходящего в экономике и повышению качества принимаемых экономическими агентами решений.

Процесс постепенного перехода к распространению статистических данных через Интернет не обошел стороной и Россию. Основы современной российской статистики были заложены до того, как Интернет в нашей стране получил широкое распространение и до начала волны разработки информационно-статистических систем в передовых странах мира. В настоящее время продолжается публикация статистических сборников привычного формата с постепенным отходом от изготовления бумажных копий в пользу их электронных образов. Этот процесс сопровождается разработкой и наполнением баз статистических данных с доступом через Интернет, которые все в большей мере привлекают внимание пользователей, хотя и не стали еще основным каналом распространения информации.

На протяжении ряда лет пользователям доступны официальный сайт Росстата, Центральная база статистических данных (ЦБСД) [8] и Единая межведомственная информационно-статистическая система (ЕМИСС) [6, 2, 7, 5], в совокупности претендующие на роль российской информационно-статистической системы. Вместе с тем имеются основания полагать, что указанные информационные ресурсы устраивают пользователей далеко не в полной мере. Примеры, свидетельствующие в пользу этой точки зрения, будут приведены ниже (см. также [3, 4]). Пока же отметим, что многие продвинутые экономисты предпочитают получать данные о российской экономике не из информационных ресурсов Росстата, а из баз данных международных организаций.

В связи с тем что в последние годы российская статистика находится на стадии перехода от использования в качестве основных каналов распространения результатов традиционных статистических сборников к информационно-статистическим системам, возникает ряд вопросов. Каковы перспективы этого переходного процесса? Какими возможностями доступа к статистическим данным будут располагать пользователи по

¹ См., например, [11, 13]. Опыт построения и публикации временных рядов исторической статистики, накопленный в США, рассмотрен в [9].

завершении перехода? Не случится ли так, что этот процесс неоправданно затянется или зайдет в тупик?

Для того чтобы пользователи смогли получить колоссальные преимущества, предоставляемые правильно организованными информационно-статистическими системами, должна быть проведена значительная работа по разработке, созданию, наполнению и сопровождению такой ИСС. Эта работа - объемна и сложна и не дает гарантии получения приемлемого результата, если потребности тех, кому предстоит работать с системой, не будут сформулированы, осмыслены и учтены. Итогом этой работы, даже при наличии достаточных ресурсов, может стать создание системы, далеко не в полной мере устраивающей пользователей.

Наивно полагать, что устраивающая пользователей система может быть создана без их активного участия. Если пользователи не возьмут на себя труд хотя бы сформулировать требования к информационно-статистической системе, они едва ли могут рассчитывать на то, что Росстат сделает для них то, о чем они не просят.

Цель настоящей статьи - сформулировать соображения автора о том, какая информационно-статистическая система нужна пользователям, решающим задачи анализа российской экономической динамики. В работе обсуждаются требования к такой системе и соображения по организации процесса ее создания, сопровождения и развития. Рассматриваются информационные структуры, на которые должна быть ориентирована ИСС, связи между ними, отражение общей логики всей совокупности статистических данных в базе данных ИСС. Делается упор на обеспечение требований полноты, избыточности, непротиворечивости, актуальности, точности. Обсуждается состав нечисловой информации в базе данных. Рассматриваются требования к функциональным возможностям системы, вопросы ее создания, сопровождения и развития. В силу того что создание ИСС имеет значительную экономико-статистическую составляющую, она не может быть создана лишь силами специалистов по информационным технологиям, поэтому обосновывается необходи-

мость организации взаимодействия разработчиков системы с ее потенциальными пользователями.

Организация данных. Статистические службы строят не отдельные показатели, а системы показателей (подробнее см. [3]), связанных отношениями иерархической подчиненности по уровню агрегирования и в территориальном разрезе. Например, строят не один индекс потребительских цен, а целую систему индексов. Помимо сводного индекса, она включает индексы по укрупненным товарным группам, агрегированием которых может быть получен сводный. Индексы цен для каждой из укрупненных товарных групп, в свою очередь, могут быть получены агрегированием индексов менее высокого уровня и так далее вплоть до индивидуальных индексов, соответствующих отдельным товарам (услугам)-представителям. Отношения между показателями этой системы могут быть отражены ориентированным древовидным графом, корневой вершине которого соответствует сводный индекс, а листьям (концевым вершинам) - индивидуальные индексы, описывающие динамику цен и тарифов отдельных товаров (услуг)-представителей. Этот граф взаимно однозначно определяется структурой корзины товаров (услуг)-представителей.

Подобные системы строят и для многих других показателей экономической динамики. Каждая из них включает один сводный индекс, множество индивидуальных и, как правило, некоторое количество групповых, то есть индексов промежуточных уровней иерархии. Каждому показателю в такой системе соответствует некоторый иерархический уровень, отражающий стадию агрегирования в структуре корзины от индивидуальных индексов к сводному.

Помимо иерархии показателей по уровню агрегирования в структуре корзины, существуют также иерархии показателей в территориальном разрезе. В системе показателей отдельные (или даже все) индексы каждого иерархического уровня агрегирования могут строиться для Российской Федерации в целом, федеральных округов и субъектов Федерации.

Помимо таких больших систем показателей, официальная статистика включает и отдельные показатели или небольшие группы показателей, не связанные отношениями иерархической подчиненности с другими. Их примерами являются индекс базовой инфляции, при построении которого стараются исключить влияние регулируемых или подверженных заметному влиянию сезонного фактора цен и тарифов на динамику всей совокупности потребительских цен, и индекс выпуска товаров и услуг по базовым видам экономической деятельности.

Показатели экономической динамики, о которых шла речь выше, представлены не отдельными числовыми величинами, а временными рядами. В дальнейшем по мере добавления новых значений для очередных периодов эти ряды становятся все более длинными. Поскольку экономика является развивающейся системой, в ней все время происходят изменения. Появляются и/или приобретают широкое распространение одни товары и услуги, тогда как другие уходят с рынка или перестают играть на нем значимую роль. Цены на одни товары и услуги растут опережающими темпами, тогда как относительные цены других снижаются. Аналогичные изменения происходят и с количествами. Рано или поздно становится очевидным, что дальнейшее использование для построения экономического индекса прежней корзины товаров-представителей и соответствующих им весов приводит к получению недостаточно репрезентативных оценок. Возникает необходимость адаптации методики к произошедшим в экономике изменениям.

С этой целью обновляют состав корзины: в нее добавляются новые товары-представители и удаляются те из прежде использовавшихся, которые утратили актуальность. Одновременно с этим уточняются веса на основе более свежей информации. Временной ряд экономический индекс продолжает строиться по измененному алгоритму. Через некоторое время методика вновь актуализируется, и так до бесконечности.

В результате длинный временной ряд экономического индекса формируется из сегментов, в рамках каждого из которых ин-

декс рассчитан по неизменному алгоритму. Процедура объединения таких сегментов в единый временной ряд называется сцеплением, а получаемый в результате временной ряд - сцепленным [12]. Ниже такие сегменты будем называть сегментами сцепления временных рядов. Через много лет после начала построения временного ряда сцепленного индекса состав корзины товаров-представителей и веса, на основе которых строятся индексы, могут значительно измениться. При этом при правильном построении временного ряда сцепленного индекса сохраняется возможность проведения долгосрочных сопоставлений экономической динамики на основе такого временного ряда.

Итак, показателям экономической динамики соответствуют не отдельные числовые величины, а временные ряды. Соответственно, и статистика, ориентированная на проведение межвременных сопоставлений, должна формироваться из временных рядов, которые должны играть роль элементарных информационных «кирпичиков».

Временной ряд представляет собой не просто объединенный во временную последовательность набор числовых величин, а информационную структуру, элементы которой связаны между собой в единое целое требованием сопоставимости. Последнее и позволяет использовать временной ряд для анализа экономической динамики. Альтернативой временному ряду в качестве базовой информационной структуры является отдельное значение экономического показателя для некоторого периода времени. В этом случае совокупности значений показателя, содержащиеся в базе данных системы, могут не образовывать полных временных рядов сопоставимых показателей. Какие-то значения могут быть пропущены, а ряды могут содержаться в базе данных не на всю их длину.

Каждому показателю может соответствовать не один, а несколько временных рядов, имеющих разные шаги по времени - месяц, квартал, год. При построении показателей с разными шагами по времени обычно обеспечивается согласованность между ними. Скажем, если индикаторы показывают изменения уровня цен между концами соседних месяцев, то годовой темп изменения цен

получается перемножением 12 месячных темпов, соответствующих этому году. Если же это показатели динамики объемов, то согласованность состоит в том, что уровень показателя за год может быть получен суммированием или осреднением уровней за соответствующие 12 месяцев.

Время от времени возникают ситуации, когда развитие большой системы показателей экономической динамики прекращается. Это происходит, например, при серьезном изменении классификатора, определяющего структуру системы показателей. Такая ситуация возникла в России при переходе от построения показателей в разрезе отраслей ОКОНХ к видам деятельности ОКВЭД, а также при переходе от классификатора ОКВЭД к ОКВЭД2. В этих случаях жизненный цикл развития системы показателей завершается, впредь она остается неизменной, а в дополнение к ней создается другая большая система показателей, начинающая свое развитие.

Развивающаяся совокупность таких систем индикаторов уровней и динамики образует всю производимую статистическую продукцию в части показателей динамики, которую ниже для краткости будем также называть официальной статистикой. Разные системы показателей статистики, не связанные между собой отношениями иерархической подчиненности (например, системы индексов потребительских цен и индексов цен производителей промышленных товаров), тем не менее не являются независимыми друг от друга, поскольку в их основе лежат единые методологические принципы. Это обеспечивает развитой статистике высокую степень концептуального единства и дает возможность проведения сопоставлений как во времени, так и между разными показателями.

На описанную архитектуру официальной статистики оказывает влияние история развития последней, административно-территориальное и политическое устройство государства, характер его экономики, участие государства в международных организациях и в интеграционных проектах и многое другое.

Таким образом, основу официальной статистики в части показателей экономической динамики составляют огромные развивающиеся системы взаимосвязанных индикаторов, к которым предъявляется большое число требований. Соответственно, организация данных в информационно-статистической системе и ее функциональные возможности должны быть ориентированы на работу не с отдельными числовыми величинами и даже не только с отдельными временными рядами, но и с большими системами показателей.

Поскольку официальную статистику составляют большие системы взаимосвязанных показателей, каждый из которых представлен одним или несколькими временными рядами, то и храниться в базе данных информационно-статистической системы они должны в виде соответствующих структур данных. Элементарной информационной структурой, наименьшим «квантом» информации, в системе должен быть временной ряд, а не набор независимо хранящихся и обрабатываемых значений показателя для разных периодов времени. Разница здесь в том, что временной ряд является единым развивающимся объектом, который никогда не исчезает и не устаревает. Набор же значений показателя для разных периодов времени является совокупностью разных объектов.

Различия между этими двумя представлениями данных можно проиллюстрировать следующим образом. При наличии пропуска в данных эта ситуация естественным образом идентифицируется в случае представления данных в виде временного ряда, но может не быть обнаруженной, если данные хранятся в виде совокупности разных объектов. Так, в ЦБСД² и ЕМИСС³ временные ряды представлены именно в виде совокупности разрозненных значений. Если данные какого-либо показателя содержат пропущенное значение, то при выгрузке временного ряда из системы в виде одномерного массива данных этот пропуск не бывает замечен – после предшествующего пропуску значения идет следующее за пропуском безо всякого указания на пропущенное значение. В резуль-

² URL: <http://cbsd.gks.ru/>

³ URL: <http://fedstat.ru/>

тате кардинально затрудняется поиск пропусков в данных - пользователь должен просмотреть одну за другой все даты, соответствующие выгруженным значениям. Скажем, индекс цен производителей сельскохозяйственной продукции по отношению к предыдущему месяцу в ЦБСД и ЕМИСС содержит пропуск данных за декабрь 2001 г. (после данных за ноябрь 2001 г. идут сразу данные за январь 2002 г.), который крайне сложно идентифицировать. Отсутствие одного из значений временного ряда эти системы не рассматривают как пропуск данных, поскольку ни с одним из объектов базы данных никакой проблемы не возникло.

Системы иерархически связанных показателей динамики в базе данных также должны быть представлены адекватными им структурами данных. Представление таких систем в виде структур данных, включающих иерархически организованные совокупности временных рядов, а не в виде отдельных временных рядов, позволяет сохранять информацию о составе, структуре и иерархической подчиненности всех показателей большой системы. Это принципиально важно для обеспечения возможности адекватного содержательного анализа таких данных. Помимо прочего, это облегчает и контроль полноты содержащейся в системе информации подобно тому, как представление данных в виде временных рядов легко позволяет находить пропущенные значения.

База данных системы должна содержать и информацию о всех когда-либо действовавших составах и структурах корзин товаров (услуг)-представителей и все действовавшие системы весов. Информация о весах, соответствующих сегментам сцепления временных рядов показателей, естественным образом может быть учтена при организации данных в виде структур, которыми оперируют современные языки программирования и системы управления базами данных и которые хорошо известны специалистам в области информационных технологий [1].

Если при построении показателей не используются досчеты (что типично для индексов цен), то эта информация обеспечивает воспроизводимость методики - агрегированием показателей некоторого уровня иерар-

хии можно получить показатель следующего более высокого уровня иерархии в пределах сегмента сцепления временного ряда. Если же при построении показателя используются досчеты (что типично для индексов объемов), то досчетная составляющая может быть представлена дополнительным индивидуальным индексом с соответствующим ему весом. При таком подходе воспроизводимость методики обеспечивается и в этом случае.

Представление информации в виде таких структур данных способно предотвратить ее утрату с течением времени, почти неизбежную при хранении ее в бумажном виде. Эта информация необходима в первую очередь самим сотрудникам статистической службы, а также продвинутой части других потребителей статистической информации. Заметим, что обеспечение доступа к разумно организованной детализированной информации будет способствовать увеличению числа пользователей, повышению их квалификации и расширению их наиболее продвинутой, наиболее требовательной части.

В системе должна быть отражена и логика всей совокупности данных официальной статистики. Наиболее естественным здесь представляется следование сложившейся в статистической службе традиции, которая отражена в «Российском статистическом ежегоднике». Использование этой же хорошо продуманной и привычной пользователям последовательности подачи материала в базе данных информационно-статистической системы, то есть обеспечение технологической преемственности в этом аспекте, упрощает освоение системы пользователями. Почему-то это простое соображение не было учтено разработчиками ныне действующих версий ЦБСД и ЕМИСС.

Временные ряды показателей экономической динамики, какими бы длинными они ни были, должны храниться в базе данных системы целиком и целиком должны быть доступными пользователям. В настоящее время встречаются случаи, когда ретроспективная часть временных рядов хранится в базе данных отдельно от их фрагментов за последние годы [3, 4]. Практика такого искусственного разрезания единых, по сути,

временных рядов с вынесением в архив ее ретроспективной части должна быть признана неприемлемой. Когда статистическая служба публикует временной ряд показателя целиком, это означает, что она берет на себя ответственность за некоторую степень сопоставимости его уровней. Когда эти же данные по частям опубликованы в разных местах и не существует официального источника, в котором они представлены целиком, это означает, что статистическая служба ответственности за их сопоставимость на себя не берет. Способ опубликования имеет значение.

Организация метаданных. Выше обсуждалось, что статистические данные в части экономической динамики состоят не из результатов сопоставлений уровней показателей для отдельных периодов времени, а из больших систем показателей, связанных между собой отношениями иерархической подчиненности. Каждый из показателей таких систем представлен временным рядом, а чаще - несколькими рядами с различными шагами по времени. Поскольку уровни и динамика показателей определяются не только характером протекания экономических процессов, но и методиками построения показателей, то для обеспечения возможности корректного анализа экономических процессов метаданные должны содержать исчерпывающие описания деталей построения всех больших систем временных рядов на всем их протяжении.

Методики построения систем показателей не остаются неизменными. Время от времени они уточняются или даже пересматриваются, то есть они подвержены эволюции. Эволюция методик обусловлена, с одной стороны, развитием экономики, то есть эволюцией объекта исследования. Необходимость адаптации к происходящим изменениям вынуждает обновлять методики. С другой стороны, эволюция методик обусловлена развитием представлений о том, как следует проводить измерения, а также изменением возможностей в этой области.

При обновлении методики, как правило, бывает невозможно уточнить все ранее сделанные оценки показателя, поскольку для

этого требуются дополнительные исходные данные, либо это связано с неприемлемыми издержками. Более того, такое обновление бывает нецелесообразным, так как методика, адекватная для одного промежутка времени, может не быть таковой для другого. В результате, как уже отмечалось ранее, разные сегменты экономических временных рядов бывают построены по несколько различающимся алгоритмам. Это особенно характерно для длинных рядов, охватывающих десятилетия.

Эволюция методик построения экономических показателей приводит к тому, что особенности динамики показателей, которые трактуются содержательно, могут быть обусловлены изменениями методов измерения и не в полной мере соответствовать динамике анализируемых параметров системы. Поэтому пользователю важно обладать информацией не только о действующей методике. Ему должна быть доступна полная временная последовательность всех когда-либо действовавших вариантов методики (включая структуру и состав корзин и веса) с информацией о том, на каком отрезке времени действовал каждый вариант, каковы различия между вариантами методики и как они влияют на уровни и динамику показателя. В противном случае пользователь будет лишен возможности отличить изменения показателя, обусловленные изменением методики, от тех, которые обусловлены существом анализируемого экономического процесса.

В настоящее время Росстат публикует либо единственный вариант методики построения показателя экономической динамики, либо несколько вариантов, действовавших в разное время. При этом не указывается, на каком промежутке времени действовал тот или иной вариант методики. Этот статичный способ опубликования методик построения экономических временных рядов определенно не ориентирован на потребности анализа экономической динамики.

Весьма полезной представляется практика приведения в приложении к тексту методики перечня существенных изменений способов формирования значений показателей за все время с начала их построения. В каче-

стве примера сошлемся на опыт Бюро статистики труда США [10], которое дает такую информацию для индексов потребительских цен с 1890 г., то есть более чем за столетний период времени. Также весьма полезно привести хронологию основных событий, повлиявших на динамику экономических процессов с начала построения соответствующих индикаторов. Для короткой, но насыщенной потрясениями истории современной России это представляется весьма актуальным. Помимо прочего, такое описание полезно тем, что объясняет логику имевших место существенных изменений методик, поскольку методы измерения динамики показателей зависят в том числе и от характера протекания экономических процессов.

Качественные метаданные необходимы не только пользователям статистических данных за пределами статистического ведомства, но и самим статистикам. Исчерпывающее описание всего процесса построения системы временных рядов (и шире – всей совокупности больших систем показателей) дает статистикам цельный взгляд на собственную деятельность и ее эволюцию с течением времени, позволяет легче находить и устранять слабые места методик, облегчает процесс вхождения в курс дела молодым специалистам, способствует лучшей подготовке студентов соответствующих специальностей. Отсутствие исчерпывающих метаданных приводит к консервации методологических проблем, то есть к сохранению невыявленных недостатков методик.

При формировании и развитии метаданных следует стремиться к соблюдению принципа преемственности текстов методик, как впрочем и статистических сборников и других статистических публикаций. Тексты методик для новых сегментов сцепления временных рядов показателей должны представлять собой усовершенствованные и актуализированные версии соответствующих материалов с предыдущего сегмента сцепления.

Неполнота опубликованных методик приводит к утрате с течением времени понимания того, как были построены соответствующие показатели и, следовательно, как нужно их содержательно интерпретировать. Это

же крайне затрудняет конструктивную обратную связь и устраняет потребителей данных официальной статистики из процесса ее совершенствования.

Подобно тому как переход от построения совокупности числовых величин к временному ряду неизбежно ставит на повестку дня вопрос обеспечения сопоставимости данных во времени и сохранения их ретроспективы, переход от статичного описания методики к полной временной последовательности всех ее действовавших вариантов поставит на повестку дня вопрос обеспечения сопоставимости результатов и сохранения полной ретроспективы методик. Базовой структурой должна быть не одна методика, а весь «поток» методик, весь их исчерпывающий временной ряд. Он жизнеспособен и полезен в отличие от статичного описания, непонятно в какой мере отражающего алгоритм построения показателя непонятно на каком промежутке времени.

При сохранении текущего положения дел, даже если показатели станут строиться и публиковаться безупречно, все равно сохранятся серьезные проблемы с их корректным использованием. Статистические данные являются полноценными лишь тогда, когда они сопровождаются адекватными метаданными, то есть когда пользователь в состоянии понять, как построен показатель и как следует интерпретировать его значения. Рано или поздно необходимо выработать стандарты подготовки метаданных и перейти к их неукоснительному соблюдению. Иначе процесс утраты понимания способов построения показателей не прекратится.

Помимо текстов методик, в системе целесообразно сохранять и значительный объем другой нечисловой информации. Полезно аккумулировать в системе все публикации статистической службы за все время ее функционирования, то есть более чем за два века. Публикации, недоступные в настоящее время в электронном виде, легко могут быть отсканированы (без распознавания текста). Каждой публикации должен соответствовать один файл в формате pdf. Следовало бы придерживаться принципа: все, однажды опубликованное статистическим ведомством, должно быть доступно пользова-

телям информационно-статистической системы. Все вновь публикуемые материалы, включая и оперативные издания, следовало бы также навсегда размещать на этом информационном ресурсе.

Функциональные возможности. Целью создания информационно-статистической системы является обеспечение доступа к данным официальной статистики, их аккумуляция и сохранение. Ценность ИСС состоит в этом и только в этом. Поэтому ключевыми должны быть функции навигации по базе данных и выгрузки (экспорта) информации для ее последующей обработки внешними по отношению к системе программными средствами. Возможности навигации и экспорта данных должны быть адекватными потребностям анализа экономической динамики.

Поскольку официальная статистика в части данных экономической динамики представлена большими системами показателей, то ИСС должна позволять выгружать такие системы и их подмножества целиком, как единые объекты, а не только как совокупности отдельных элементов. Данные всей большой системы показателей или ее подмножества должны выгружаться в результате выполнения одного запроса и с сохранением их внутренней логики, а не в виде совокупности разрозненных временных рядов или скалярных величин.

Как отмечалось выше, структура большой системы показателей в пределах одного сегмента сцепления временных рядов определяется ориентированным древовидным графом. Для разных сегментов сцепления структуры могут несколько различаться, главным образом за счет отличий в составах товаров-представителей и групповых индексов нижних уровней агрегирования. Исходя из потребностей задач анализа экономической динамики, представляется целесообразной реализация следующих возможностей выгрузки данных. Должна существовать возможность выгрузки данных, соответствующих подграфу, определяемому вершиной произвольного уровня иерархии и заданным числом менее высоких уровней. Запрос на выгрузку данных, помимо указа-

ния на корневую вершину подграфа и уровень его концевых вершин, должен включать также указания на территориальную единицу, шаг по времени выгружаемых временных рядов (месяц, квартал, год) и границы промежутка времени, данные для которого подлежат выгрузке. Заметим, что этот подход позволяет выгружать одной командой, помимо прочего, как данные всей большой системы показателей для заданной территориальной единицы, так и отдельные временные ряды.

Если на промежутке времени, данные которого подлежат выгрузке, структура выгружаемого подмножества данных не претерпевала изменений, то экспорт данных может состоять в формировании файла двумерной электронной таблицы, одно из измерений которой соответствует оси времени, а второе - показателям. При этом должны выгружаться уникальные коды показателей, их наименования, территориальная принадлежность, указания на уровень иерархии в системе и веса, с которыми показатели данного уровня агрегируются при построении показателя следующего более высокого уровня иерархии. Такое представление данных позволяет с минимальными для пользователя издержками существенно расширить его аналитические возможности. Скажем, можно будет легко анализировать влияние изменений индивидуальных индексов на изменение агрегированного или строить агрегированный индекс по любому подмножеству совокупности индивидуальных.

Если же на промежутке времени, данные которого подлежат выгрузке, структура выгружаемого подмножества данных изменилась, то должны экспортироваться несколько электронных таблиц (или несколько листов одной электронной таблицы), каждая из которых содержит данные на очередном интервале времени, в пределах которого структура выгружаемого подмножества неизменна. Структура каждой электронной таблицы в этом случае должна быть той же, что и в предыдущем. Заметим, что при таком подходе запрос на выгрузку отдельного временного ряда показателя любого уровня агрегирования позволяет выгрузить ряд максимально возможной длины даже и в том слу-

чае, если этот ряд принадлежит сегментам сцепления с разным составом индексов.

Помимо описанной возможности выгрузки для заданной территориальной единицы иерархии показателей по уровню агрегирования, целесообразно также обеспечить аналогичную возможность выгрузки заданного показателя для иерархии территориальных единиц.

Представляется нецелесообразным предоставление возможности указания отдельных месяцев, для которых пользователь хотел бы выгрузить данные временных рядов из системы. Кому и для чего может потребоваться временной ряд без апрельских и июльских значений? А если такой ряд и потребуется, то он тривиально может быть получен самим пользователем на основе выгруженного из системы полного ряда.

Для обеспечения удобного и эффективного доступа к информации система должна предоставлять адекватные средства навигации по базе данных. Поскольку официальная статистика содержит огромное количество показателей, весьма полезным представляется присвоение каждому показателю уникального кода. Кодировка показателей позволяет избежать путаницы, неизбежной при словесной идентификации показателей, а также облегчает создание баз данных и поиск показателей в них. Отсутствие уникальных кодов показателей весьма затрудняет навигацию в ЦБСД и ЕМИСС. Реализованная в настоящее время функция поиска показателя по его полному наименованию не всегда эффективна, поскольку наименования одного и того же показателя в разных местах базы данных могут несколько различаться, а также потому, что пользователь должен знать точные наименования необходимых ему показателей, то есть держать в голове все необходимые классификаторы (подробнее см. [3]). В дополнение к поиску по полному наименованию показателя представляется целесообразной реализация контекстного поиска по ключевым словам и фрагментам полного наименования.

Помимо навигации по кодам показателей и контекстного поиска по ключевым словам, система должна предоставлять возможность навигации и по содержанию базы

данных. При этом навигация должна быть основана не на полных классификаторах, а на их подмножествах, соответствующих фактически содержащейся в базе данных информации. Так, классификатор единиц измерения содержит большое число элементов, тогда как размещенные в базе данных показатели статистики могут использовать лишь несколько из них. Если в основу алгоритма навигации будет положена необходимость указания элемента из полного классификатора, то такая навигация будет предельно неудобной, поскольку большинству элементов классификатора в базе данных не будет соответствовать никакой информации. Именно так реализована функция навигации в ЦБСД.

При реализации функции поиска по содержанию базы данных показатели и территориальные единицы должны быть разумно упорядочены. Для показателей это означает, что первым в списке должен быть сводный показатель, а показатели каждого следующего, менее высокого уровня агрегирования должны перечисляться непосредственно вслед за соответствующим им агрегированным показателем и в последовательности, соответствующей использованной в методике (скажем, в порядке убывания их значимости). Для территориальных единиц это означает, что первой в списке должна быть Российская Федерация, затем первый из устоявшейся последовательности перечисления федеральных округов, затем соответствующие ему субъекты Федерации также в устоявшейся последовательности, затем следующий федеральный округ и т. д. В настоящее время в ЦБСД и ЕМИСС элементы списков либо упорядочены по алфавиту, либо - произвольным образом. В любом случае - без выделения уровней иерархии и отношений подчиненности.

Для обеспечения удобной навигации важно задать в системе осмысленные значения параметров умолчания. Скажем, разумным параметром умолчания в иерархии индексов потребительских цен представляется сводный ИПЦ по всем товарам и платным услугам, в иерархии территориальных единиц - Российская Федерация в целом и т. п. В ныне действующих версиях ЦБСД и ЕМИСС во многих случаях территории, группы и

виды товаров и услуг, интервалы времени, виды данных установлены по умолчанию без всякой логики, как попало. Это весьма затрудняет работу с системами. Очевидно, что разработчикам систем содержащиеся в них показатели не говорили решительно ничего.

Функции навигации и выгрузки данных должны быть продуманы и реализованы с максимально возможной тщательностью. В то же время едва ли целесообразно стремиться к расширению состава функций системы, не связанных непосредственно с доступом к данным. Так, вполне можно обойтись без блока визуализации данных в графическом виде. Во-первых, ресурсы, выделяемые на создание системы, всегда являются ограниченными. Поэтому имело бы смысл сконцентрировать их на совершенствовании возможностей доступа к данным и на повышении объема и качества самих данных, то есть инвестировать их в развитие ключевых характеристик системы. Никакие графические и прочие оформительские возможности не способны компенсировать отсутствия данных в системе или неполноту функциональных возможностей работы с ними (скажем, невозможность выгрузить из базы данных всю систему показателей целиком с сохранением информации об ее структуре). Во-вторых, пользователям доступна масса программных пакетов, позволяющих строить графики на любой вкус. Поэтому нет необходимости включать графические возможности в систему. Все равно они не покроют всех потребностей разных категорий пользователей.

Это же относится и к другим дорогостоящим блокам (скажем, к геоинформационной системе). Если без них можно обойтись, то так и стоило бы сделать (во всяком случае, в первых версиях системы). Начинать нужно явно не с этого, а с того, чтобы довести до приемлемого уровня те возможности, без которых нельзя обойтись. Безусловно, необходимы лишь статистические данные, возможности навигации и средства выгрузки их из системы в форматах, ориентированных на дальнейшую электронную обработку.

Чем проще будет система и чем меньшего объема ресурсов она будет требовать для своего создания и эксплуатации, тем мень-

ше ее судьба будет зависеть от переменчивой конъюнктуры. Длительность (или даже неограниченность) жизненного цикла обсуждаемого информационного ресурса требует учитывать требование его переносимости на системные средства, которые могут быть разработаны в будущем. Поскольку невозможно уверенно прогнозировать изменения, которые могут произойти хотя бы в ближайшие десятилетия в операционных системах, системах управления базами данных, пакетах обработки статистических данных, системах научной графики, то при создании ресурса необходимо использовать минимальный набор наиболее стандартных системных средств и максимально упростить и стандартизировать интерфейсы пользователя и обмена данными.

Система должна быть самодостаточной: пользователь должен иметь возможность работать с ней без привлечения дополнительной информации. Система должна быть «прозрачной» для пользователя, понятной ему на интуитивном уровне. Она должна быть аскетичной и консервативной.

Временные ряды показателей высокой частоты (квартальные и месячные) обычно бывают не вполне пригодными для проведения анализа экономической динамики непосредственно, то есть без проведения дополнительной обработки. Такая обработка обычно предполагает проведение сезонной корректировки. Соответственно, необходимо определиться с отношением информационно-статистической системы к проведению сезонной корректировки.

Представляется, что ИСС не должна предоставлять пользователям возможности проведения сезонной корректировки по тем же причинам, по которым нецелесообразно предоставлять им графические возможности. Вместе с тем, если статистическая служба проводит сезонную корректировку показателей, то ее результаты могут быть загружены в базу данных и стать доступными пользователям, наряду с данными, не подвергнутыми корректировке. Это не может создать проблем, поскольку пользователю будут доступны как исходные (нескорректированные) данные, так и результаты обработки, проведенной специалистами статистической

службы. Если эти результаты не устраивают пользователя, то он может провести сезонную корректировку самостоятельно, выгрузив исходные данные из ИСС и загрузив их в пакет декомпозиции экономических временных рядов. Для этого принципиальной является лишь возможность получения пользователем данных в формате, ориентированном на дальнейшую электронную обработку.

Организация процессов создания, сопровождения и развития. Создание больших программных систем принято начинать с анализа требований к ним. На этом этапе разработки необходимо, в частности, понять, какие категории специалистов станут пользователями системы и какие задачи они будут решать с ее помощью. То есть необходимо определить целевую аудиторию системы и ее назначение.

В дальнейшем следует придерживаться принципа пошаговой детализации. В соответствии с ним сначала нужно сформулировать первый вариант требований к системе и согласовать их с предполагаемыми пользователями. На основе полученного в результате такого обсуждения набора требований необходимо разработать внешние спецификации системы, то есть определить, как она будет выглядеть для конечного пользователя. Это нужно вновь обсудить с предполагаемыми пользователями, вернувшись при необходимости к предыдущему этапу с целью уточнения требований к системе. После стабилизации, в результате серии таких обсуждений, набора требований к системе и ее внешних спецификаций, можно переходить к ее проектированию и программной реализации, после чего полученный вариант системы следует вновь обсудить с пользователями, возвращаясь при необходимости к уточнению результатов предыдущих этапов.

Если разработку системы вести в иной последовательности (скажем, если сначала создать полностью функционирующий вариант системы и лишь после этого анализировать реакцию пользователей), то получим длительный, мучительный, дорогостоящий процесс создания все новых и новых версий

системы, каждой из которых многие пользователи будут недовольны.

Принципиально важным представляется подключение к процессу создания системы на самых ранних его этапах не только специалистов в области информационных технологий, но и производителей данных экономической статистики и ее потребителей, то есть тех специалистов, которым предстоит с ней работать. Среди представителей потребителей статистической информации, которых имело бы смысл привлечь к выработке требований к системе и определению ее облика, должны быть сотрудники основных ведомств экономического блока системы государственного управления, работники аналитических структур, осуществляющих экспертную поддержку процесса выработки экономической политики, представители академического сообщества, преподаватели, журналисты. Привлекать нужно не администраторов и «свадебных генералов», а активно работающих с данными экономической статистики специалистов, кровно заинтересованных в получении инструмента, позволяющего повысить качество и снизить издержки их работы.

Если процесс создания информационно-статистической системы будет отдан на откуп специалистам по информационным технологиям, то едва ли можно рассчитывать на достижение приемлемого результата. Какими бы квалифицированными в своей области они ни были, заменить производителей данных экономической статистики и их потребителей при определении набора требований к системе и формировании ее облика они никогда не смогут. Деятельность по разработке информационно-статистической системы включает существенную экономико-статистическую составляющую и не может быть сведена лишь к задаче из области информационных технологий. Достижение приемлемого результата едва ли возможно и в том случае, если со стороны потенциальных пользователей системы будут учтены требования лишь производителей статистических данных. Хотя учет потребностей этой части сообщества пользователей совершенно необходим, сформировать требования внешних по отношению к статисти-

ческому ведомству потребителей информации они едва ли смогут. Причина в том, что сотрудники Росстата не вовлечены непосредственно в аналитическую и исследовательскую деятельность.

Создание и развитие информационно-статистической системы следует рассматривать не как проект, а как процесс. Это - не разовое мероприятие, а непрерывная деятельность, которая, начавшись, не может быть прекращена, хотя ее интенсивность может изменяться в весьма широких пределах. Для организации этой деятельности в составе статистической службы целесообразно создать постоянно действующую группу специалистов, которая будет координировать процесс формирования требований к системе, организовывать взаимодействие с ее потенциальными пользователями, разрабатывать технические задания для разработчиков системы, осуществлять контроль хода выполнения работ и их приемку. Функции данной группы напоминают функции отдела капитального строительства, который имеется в структуре многих крупных организаций и который занимается обеспечением процесса строительства, непосредственно не участвуя в нем. Обсуждаемая группа может, в частности, организовывать взаимодействие с сообществом потребителей статистической информации, готовить материалы к обсуждениям, направлять их на экспертизу и т. п.

Процесс развития системы невозможен без организации эффективного канала обратной связи. Его задачи - исправление выявляемых пользователями ошибок в информации и в работе системы, а также учет пожеланий пользователей по развитию системы. Обеспечение функционирования такого канала должно быть одной из задач обсуждаемой группы.

Информационно-статистическую систему недостаточно лишь разработать. Необходимо обеспечить наполнение ее базы данных и организовать процесс последующего обновления содержимого базы данных и сопровождения системы. При этом следует соблюдать стандартные требования, предъявляемые к базам данных, такие, как полнота, избыточность, непротиворечивость, точность, актуальность.

Под полнотой представленной в базе данных информации здесь понимается хранение в системе полного набора когда-либо существовавших систем показателей на максимально возможную глубину ретроспективы каждого показателя и с максимально возможной их детализацией. Все, что не противоречит законодательству, регулирующему статистическую деятельность, в соответствии с требованием полноты должно содержаться в базе данных ИСС и быть доступно ее пользователям.

В соответствии с требованием избыточности значение каждого показателя должно храниться лишь в одном месте. Не должно быть ситуаций, когда значение одного и того же показателя введено в систему неоднократно. Скажем, нельзя допускать, чтобы значение одного экономического индекса за некоторый год независимо содержалось в двух местах базы данных.

Соблюдение требования избыточности способствует обеспечению выполнения требования непротиворечивости данных, в соответствии с которым, в частности, не должно быть возможности получения пользователем при обработке его разных запросов различающихся значений одного и того же показателя. Такая ситуация возможна, например, если значения показателя были помещены в два разных места в базе данных, в одном из которых они впоследствии были уточнены. Либо если значения показателя помещены в разные места базы данных с разной точностью.

Представленные в базе данных показатели должны удовлетворять требованию точности в двух аспектах. Во-первых, они должны содержать необходимое число десятичных разрядов. В настоящее время распространена практика опубликования данных экономической динамики с недостаточной точностью. Скажем, если цены растут на 4% за год, то представление соответствующих изменений за месяц с точностью до одного знака после запятой представляется недостаточным. Во-вторых, требование точности подразумевает организацию эффективного процесса исправления ошибок. Известно, что всякий достаточно обширный массив информации обязательно содержит некоторое

количество ошибок. Впоследствии в процессе работы с данными эти ошибки могут быть идентифицированы и исправлены. Если этого не делать, то база данных информационно-статистической системы постепенно превратится в «электронную помойку», ценность которой невелика.

Наконец, данные должны удовлетворять требованию актуальности, в соответствии с которым при появлении информации за очередные периоды времени, а также при уточнении предварительных и ретроспективных значений показателей это немедленно должно отражаться в базе данных. Поскольку в России основным каналом распространения данных традиционно являются печатные издания (или электронные документы, по содержанию полностью соответствующие печатным изданиям), то базы данных порой рассматриваются в качестве вторичного (не основного) канала распространения данных, обеспечению актуализации которых не уделяется достаточного внимания. Это значительно снижает ценность таких баз данных. Именно ИСС должна рассматриваться в качестве основного канала распространения данных. Регламентация процесса обновления данных в ИСС не должна предполагать привязки к срокам выхода каких бы то ни было статистических сборников или печатных изданий.

Обсуждение. Выше приведены соображения о том, какой хотел бы видеть российскую информационно-статистическую систему пользователь, решающий задачи анализа экономической динамики. Многие из этих соображений являются очевидными следствиями потребностей соответствующих задач. В связи с этим интересно понять, в какой мере потребности задач анализа экономической динамики учтены в действующих версиях ЦБСД и ЕМИСС.

Знакомство с функциональными возможностями и содержимым ЦБСД и ЕМИСС показывает, что они не вполне ориентированы на работу с временными рядами, то есть с базовыми информационными структурами, адекватными задачам анализа экономической динамики (подробнее см. [3]). Данные в этих системах представлены в виде совокуп-

ности отдельных числовых величин, а не в виде временных рядов. При выгрузке данных из ЕМИСС необходимо указывать не начало и конец интервала времени, а перечень отдельных лет и месяцев. Многие временные ряды охватывают не всю историю построения соответствующего показателя. Часть рядов разрезана произвольным образом с вынесением ретроспективной части в архив. Данные приведены в системах с разной степенью детализации в разрезе показателей и территорий и за разные интервалы времени.

Некоторые данные представлены только показателями по отношению к соответствующему месяцу или кварталу предыдущего года, что не позволяет использовать их для проведения полноценного анализа краткосрочных тенденций. Данные в разных цепных формах (по отношению к предыдущему месяцу, по отношению к соответствующему месяцу предыдущего года) не всегда согласуются друг с другом. Большие системы показателей не могут быть выгружены из ЦБСД и ЕМИСС целиком с сохранением их внутренней логики, а лишь отдельными показателями (ЦБСД), либо целиком, но в алфавитном порядке без указания на иерархический уровень (ЕМИСС), то есть без информации о взаимной подчиненности. Навигация по базам данных реализована неудобно. Так, выбирать признак показателя в ЦБСД приходится из полного классификатора, а не из его подмножества, для которого в базе данных имеется информация. Параметры умолчания заданы без всякой логики. Метаданные малоинформативны. Требования полноты, избыточности, непротиворечивости, актуальности нарушены.

Анализируя действующие версии ЦБСД и ЕМИСС трудно понять, на решение каких задач они ориентированы. Определенно просматривается лишь ориентация на построение простых таблиц и графиков, что может быть полезным при подготовке статистических сборников и периодических изданий. Эти возможности должны быть востребованы сотрудниками статистического ведомства, однако потребности внешних по отношению к Росстату пользователей неизмеримо шире.

Вызывает беспокойство то обстоятельство, что при наличии очевидных проблем с

информационно-статистическими ресурсами не заметно признаков активности по части их решения. На заседаниях Научно-методологического совета Росстата соответствующие вопросы не рассматривались. В последние годы эти проблемы не обсуждались и на страницах журнала «Вопросы статистики». Состоялось лишь обсуждение Концепции модернизации официального сайта на заседании Общественного совета Росстата 11 мая 2016 г. Из размещенных на сайте материалов этого заседания⁴ узнаем, что речь идет о совершенствовании технической стороны сайта, а по части содержательных вопросов предполагается лишь «сохранить текущую функциональность, накопленную информацию и разработанные сервисы, обеспечить преемственность». Другими словами, решать имеющиеся проблемы доступа к данным не предполагается. Похоже, развитие информационно-статистических ресурсов в Росстате производится лишь силами специалистов в области информационных технологий без полноценного участия тех, кому с ними предстоит работать. Этим можно объяснить отсутствие учета потребностей внешних по отношению к статистическому ведомству пользователей.

Приведение данных и метаданных в соответствие с описанными выше (или какими бы то ни было иными, учитывающими потребности пользователей) требованиями не сводится лишь к организации существующих массивов статистической информации в базе данных, а предъявляет дополнительные требования и к самим данным и метаданным. Другими словами, эта деятельность помимо составляющей по части информационных технологий имеет и существенную экономико-статистическую составляющую. Проблемы существующих систем заключаются как в программных средствах и организации базы данных, так и в качестве содержимого. Поэтому требуется не только разработка системных средств работы с данными и наполнение базы данных существующей информацией, но и проведение масштабной деятельности по совершенствованию, упорядочению практики построения данных и метаданных. Создание правильно организо-

ванной информационно-статистической системы способно заметно повлиять на работу Росстата.

Отсутствие полноценного по современным меркам канала распространения информации, делая интерфейс российской статистики узким местом, снижает эффективность ее функционирования. Представляется, что создание качественного интерфейса официальной статистики невозможно без скоординированных совместных усилий специалистов в области информационных технологий с производителями и потребителями статистической информации. В противном случае существует риск того, что российская статистика в какой-то мере будет функционировать вхолостую.

Литература

1. Ахо А.В., Хопкрофт Дж.Э., Ульман Дж.Д. Структуры данных и алгоритмы. М.: Вильямс, 2016. 400 с.
2. Белов Н.В. Росстат как координатор ЕМИСС // Вопросы статистики. 2011. № 10. С. 51-53.
3. Бессонов В. Что сохранит для истории современная российская статистика? // Вопросы экономики. 2015. № 1. С. 125-146.
4. Бессонов В.А. Взгляд на российскую статистику со стороны пользователя // Вопросы статистики. 2009. № 5. С. 50-61.
5. Забелин В.Н. ЕМИСС ответит на любой вопрос // Вопросы статистики. 2012. № 9. С. 22-24.
6. Концепция развития ИВС Росстата. М.: Росстат, 2010. 57 с.
7. Пашинцева Н.И. О создании Единой межведомственной информационно-статистической системы // Информационные ресурсы России. 2011. № 1. С. 7-8.
8. Пряхина Е.В., Осипенко Т.М., Аксенова Т.В. Центральная база статистических данных Росстата // Вопросы статистики. 2005. № 3. С. 24-27.
9. Симчера В.М., Соколин В.Л. Зарубежный опыт построения и публикации долговременных историко-статистических рядов // Вопросы статистики. 2000. № 11. С. 3-14.
10. BLS Handbook of Methods. Chapter 17. The Consumer Price Index. Bureau of Labor Statistics. June 2015. 107 p. URL: <http://www.bls.gov/opub/hom/pdf/homch17.pdf>
11. Carter S.B., Gartner S.S., Haines M.R., Olmstead A.L., Sutch R., Wright G. (eds.) Historical Statistics of the United States. Millennial Edition. N.Y.: Cambridge University Press, 2006. 5 volumes. 4500 est. p.
12. Forsyth F.G., Fowler R.F. The Theory and Practice of Chain Price Index Numbers // Journal of the Royal Statistical Society. Ser. A. 1981. Vol. 144. Part. 2. P. 224-246.
13. Mitchell B.R. British Historical Statistics. Cambridge University Press, 2011. 898 p.

⁴ URL: http://www.gks.ru/free_doc/new_site/rosstat/os/press-reliz-os.pdf

WHAT THE RUSSIAN INFORMATION AND STATISTICAL SYSTEM SHOULD BE LIKE?

Vladimir A. Bessonov

Author affiliation: National Research University Higher School of Economics (Moscow, Russia). E-mail: bessonov@hse.ru

Information and communication technology has reached a level of development that official sites of statistical offices and information resources stored on them has become the major channel for the distribution of statistical data. Therefore issues related to dissemination of statistical information should now be discussed in the context of internet-based databases, i.e. information and statistical systems. Transition from statistical handbooks and publications to such systems can lead to a radical increase in the volume and quality boost of statistical information made available to the users along with drastic reduction in data access costs.

For several years now users have been able to access the official Rosstat website, the Central Statistical Database and the Unified Interdepartmental Statistical Information System that combined claim to the role of Russian information and statistical system. However it is safe to say that the users of all the abovementioned resources are not exactly content and satisfied with them.

The purpose of this article is to organize author's thoughts on what kind of information and statistical system will suit the users tasked with analyzing Russian economic dynamics more.

The article discusses what's required of such system and puts forward suggestions for structuring the process of its organization, maintenance and development. The author reviews information structures that this system should look up to, relationships between them, general logic behind the total data set in the database. Emphasis is being placed on ensuring the following requirements: completeness, non-redundancy, consistency, actuality, accuracy and precision. The composition of non-numeric information in the database is discussed. The article examines system functionality, issues related to its establishment, maintenance and development. Because establishment of this system has a substantial economic and statistical component, it can not be done only by the efforts of IT specialists. It is necessary to organize interaction between system developers and its potential users.

Keywords: database, time series, data access, economic statistics, index number.

JEL: C82, C88.

References

1. Aho A.V., Ullman J.D., Hopcroft J.E. Data Structures and Algorithms. Reading, Mass.: Addison-Wesley, 1987. 427 p. (Russ. ed.: Akho A.V., Khopkroft Dzh.E., Ul'man Dzh.D. Struktury dannykh i algoritmy. M.: Vil'yams, 2016. 400 s.)
2. Belov N.V. Rosstat kak koordinator EMISS [Federal State Statistics Service as a coordinator of the Unified Interdepartmental Information and Statistical System]. *Voprosy statistiki*, 2011, no. 10. pp. 51-53. (In Russ.).
3. Bessonov V. Chto sokhranit dlya istorii sovremennaya rossiyskaya statistika? [What Will the Modern Russian Statistics Have in Store for History?] *Voprosy ekonomiki*, 2015, no. 1, pp. 125-146. (In Russ.).
4. Bessonov V.A. Vzgl'yad na rossiyskuyu statistiku so storony pol'zovatelya [User's onlooking at Russian statistics]. *Voprosy statistiki*, 2009, no. 5, pp. 50-61. (In Russ.).
5. Zabelin V.N. EMISS otvetit na lyuboy vopros [UniSIS will answer any question]. *Voprosy statistiki*, 2012, no. 9, pp. 22-24. (In Russ.).
6. Kontseptsiya razvitiya IVS Rosstata [The Concept of Development of the Rosstat Information and Computing System]. Moscow, Federal State Statistics Service, 2010/ 57 p. (In Russ.).
7. Pashintseva N.I. O sozdaniy Yedinoy mezhvedomstvennoy informatsionno-statisticheskoy sistemy (EMISS) [On the Creation of a Single Interdepartmental Information and Statistical System]. *Informatsionnyye resursy Rossii*, 2011, no. 1, pp. 7-8. (In Russ.).
8. Pryakhina Ye.V., Osipenko T.M., Aksenova T.V. Tsentral'naya baza statisticheskikh dannykh Rosstata [The Central statistical data base of Rosstat]. *Voprosy statistiki*, 2005, no. 3, pp. 24-27. (In Russ.).
9. Simchera V.M., Sokolin V.L. Zarubezhnyy opyt postroyeniya i publikatsii dolgovremennykh istoriko-statisticheskikh ryadov [Foreign experience of estimation and publication of long-term historical statistical data series]. *Voprosy statistiki*, 2000, no. 11, pp. 3-14. (In Russ.).
10. BLS Handbook of Methods. Chapter 17. The Consumer Price Index. Bureau of Labor Statistics. June 2015. 107 p. Available at: <http://www.bls.gov/opub/hom/pdf/homch17.pdf>.
11. Carter S.B., Gartner S.S., Haines M.R., Olmstead A.L., Sutch R., Wright G. (eds.) Historical Statistics of the United States. Millennial Edition. N.Y.: Cambridge University Press, 2006. 5 volumes. 4500 est. p.
12. Forsyth F.G., Fowler R.F. The Theory and Practice of Chain Price Index Numbers. *Journal of the Royal Statistical Society*, ser. A. 1981, vol. 144, part. 2. P. 224-246.
13. Mitchell B.R. British Historical Statistics. *Cambridge University Press*, 2011. 898 p.