

УИС РОССИЯ: Социально-экономическая статистика по муниципалитетам РФ * в формате реляционной базы

© Богомолова А.В.^{2,3)}, Карасев О.И.^{2,3)}, Сеннов Р.^{2,4)}, Юдина Т.Н.^{1,2)}

¹⁾ Научно-исследовательский вычислительный центр МГУ им. М.В.Ломоносова

²⁾ АНО Центр информационных исследований

³⁾ Экономический факультет МГУ им. М.В. Ломоносова

⁴⁾ Физический факультет МГУ им. М.В. Ломоносова

yudina@mail.cir.ru

Аннотация

Публикация посвящена описанию баз данных «Социально-экономическая статистика по муниципалитетам РФ» и «Бюджетная статистика по муниципалитетам РФ» в составе Университетской информационной системы РОССИЯ (www.cir.ru) Рассмотрена концепция баз данных, их содержание, особенности применяемых технологий и элементы академического сервиса.

Оба ресурса представлены в формате реляционной базы, формат поддерживает исследование каждого показателя или группы выбранных показателей в динамике за указанный период времени в ретроспективе до 10 лет и в разрезе страна-регион-муниципалитет. Интеграция баз данных с документами других 60 коллекций и технологические решения УИС РОССИЯ позволяю проводить междисциплинарные исследования с использованием компьютерных методов анализа и применять математические приемы обработки больших объемов данных и документов. Расширяется научный потенциал специалистов - географический анализ дает возможность рассмотреть каждый муниципалитет в сравнении с соседними и на фоне социальных процессов в регионе и в стране, сравнивать динамику развития муниципалитетов и регионов, проводить системный и т.н. «панорамный» анализ. Содержательное наполнение, технологические решения, регулярное обновление ресурсов, пользовательские сервисы в комплексе формируют среду для изучения российского общества на уровне

реальной жизни людей. Такие исследования имеют не только научную ценность, но и практическую значимость для органов управления всех уровней.

1. Статистика на уровне муниципальных образований для исследований и как информационный ресурс для органов власти и местного самоуправления

Глубина статистических измерений – один из показателей информационного развития общества. Во всех развитых странах социально-экономическая статистика собирается в разрезе начальных территорий. В США и большинстве государств Европы поддерживается статистика по домохозяйствам, интегрированная в сводную информационно-статистическую систему страны. В РФ отсутствует единый ресурс, содержащий систему показателей социально-экономического развития в разрезе «страна – федеральный округ – субъект федерации – муниципалитет». В ежегодных статистических публикациях Федеральной службы государственной статистики РФ (Госкомстата России) представлены данные преимущественно общедоменного и регионального уровней. Практически единственный комплексный материал о социально-экономическом развитии муниципалитетов, издаваемый Госкомстатом России, - база данных «Паспорта городов», содержит данные по 1051 муниципалитету, в основном, по крупным городам (столицы субъектов Федерации, города с численностью населения более 100 тыс. человек). В некоторых других сборниках Госкомстата России показатели по муниципалитетам представлены выборочно, причем, по малым городам объем данных незначителен, по большинству муниципальных образований данные в централизованном порядке не собираются.

Даже представленные в «Паспортах городов» данные требуют экспертизы на предмет сопоставимости показателей по разным муниципалитетам, поскольку одни и те же показатели зачастую имеют разную трактовку и методологию исчисления. [1]. Формирование полноценной базы по местной статистике не предусматривается и в рамках государственной программы «Электронная Россия», по крайней мере, до 2010 года. Отсутствие данных в разрезе начальных территориальных единиц – серьезная информационная проблема, затрудняющая проведение полноценных исследований российского общества на уровне реальной жизни граждан, особенно живущих в сельской местности, в удаленных и северных районах. К проведению таких исследований готов целый ряд научных коллективов в Москве, Санкт Петербурге, Новосибирске и многих других городах России, способных проводить системные исследования с использованием математических методов обработки статистических данных. Востребованность ресурса очевидна и со стороны органов управления страной – комплексный социальный анализ должен предшествовать принятию управленческих решений на всех уровнях власти.

Представляется рациональным и технологичным разработка такого ресурса на базе уже действующей реляционной базы по социально-экономическим показателям в разрезе страны и регионов, созданной в рамках Университетской информационной системы РОССИЯ (УИС РОССИЯ, www.cir.ru). УИС РОССИЯ включает два предметно-ориентированных ресурса – базы данных «Статистика России» и «Бюджетная система РФ» [3], описанные в настоящей статье работы планируются коллективом как следующий этап развития обоих ресурсов и УИС РОССИЯ в целом.

2. База данных по социально-экономической статистике муниципалитетов

Текущая версия БД «Статистика России» интегрирует статистические массивы и публикации по социально-экономическому развитию страны, издаваемые правительственными учреждениями, исследовательскими центрами, материалы научных журналов и формирует сводный информационный ресурс для мониторинга социальных и экономических процессов в России на уровне страны и регионов. База содержит 95 ежегодных публикаций (сборников) Госкомстата России, включает статистические данные и аналитические доклады, в том числе ежегодные публикации «Российский статистический ежегодник», «Регионы России», «Промышленность России» и ряд других отраслевых сборников, а также «Социальное положение и уровень жизни населения в России», «Демографический ежегодник», «Россия и страны мира», «Россия и страны СНГ»,

«Население России за 100 лет (1897–1997 гг.)», «Национальные счета России», «Финансы России», «Цены в России» и другие сборники. Ретроспектива изданий – с 1996 года. Блок ежемесячных публикаций представлен статистическими таблицами «Краткосрочные экономические показатели Российской Федерации» в ретроспективе с 1999 года. Представлен раздел методологического сопровождения статистических данных, реализованный с использованием публикации Госкомстата России «Методологические положения по статистике» и выполненный как многоуровневый электронный справочник с развитыми гипертекстовыми связями. Краткие пояснения составлены с использованием материалов соответствующих статистических сборников и привязаны к конкретным таблицам. Развернутые методологические положения детально описывают стандарты российского статистического учета. Имеется гипертекстовый Глоссарий статистических терминов. Блок аналитических материалов содержит официальные издания – ежемесячный доклад «Социально-экономическое положение России» Госкомстата России, Мониторинг Министерства экономического развития и торговли РФ, Вестник Банка России и публикации независимых исследовательских центров, среди которых – Российско-Европейский центр экономической политики, Фонд «Содействие построению и развитию Индекса Потребительских Настроений», Бюро экономического анализа, Независимый институт социальной политики (доклады по программе «Социальная политика») и другие. Объем статистической части базы – свыше 50 000 таблиц. Статистические данные могут быть получены и в формате *MS Excel 97* и импортированы в программы статистического анализа (*SPSS, Statgraphics, Statistica*) и пакеты табличного анализа (*MS Excel, Lotus 1-2-3* и др.). С сентября 2003 года статистическая коллекция сборника Госкомстата России «Регионы России» доступна в формате реляционной базы «Интерактивная статистика». База поддерживает поиск значения конкретного показателя (или выборки показателей) по состоянию на определенную дату/период времени и с конкретной территориальной привязкой. Интерфейс позволяет осуществлять построение сводных аналитических таблиц на основе пользовательского запроса. Систематизация данных осуществляется путем применения единого каталога статистических показателей. Пользователь может выбрать несколько показателей, в том числе полученных из разных источников, которые будут сведены в единую таблицу для определенного периода времени и по сопоставимому кругу регионов. Реализована функция визуализации данных социально-экономической статистики в форме картограмм. Модуль отображения данных на карте интегрирован с реляционной базой, что предоставляет пользователю возможность

оперативно вывести в форме картограммы любой показатель из сводной таблицы, построенной в результате запроса.

Повторим, что в текущей версии системы поддерживаются показатели на уровне страны в целом и регионов. Несмотря на отсутствие централизованного процесса сбора статистики по муниципалитетам, данные на уровне муниципальных образований собирались и собираются некоторыми федеральными министерствами, научными институтами и исследовательскими центрами, отдельными специалистами. В последние годы в стране выполняются проекты, включающие, в том числе, сбор муниципальной социально-экономической статистики для конкретных исследовательских задач. При отсутствии единых требований, каждый коллектив формирует ресурсы с учетом для своих внутренних нужд, конкретных задач, зачастую коллекции состоят из разрозненных массивов, подготовленных разными подразделениями, не полны и представлены с разной ретроспективой, выборочным обновлением. Сведение данных из разных источников в единую базу и формирование полноценного ресурса – трудоемкий процесс, поскольку предполагает решение комплекса организационных, технических, методологических вопросов, в том числе, легальное получение ресурсов от правообладателей, обследование и проверку на полноту коллекций, сбор недостающих первичных данных. Одновременно должны быть решены и технологические задачи, к примеру, разработана система классификации показателей, созданы программ для предварительной обработки источников. Практика показывает, что даже данные сборников Госкомстата России требуют экспертизы и значительной дополнительной работы для сведения в базу, так как показатели опубликованы в разных тематических сборниках, сгруппированы в разные по своей структуре таблицы; причем таблицы с одинаковой структурой, полученные из разных первоисточников; могут содержать данные, методология расчета которых различается и т.п. Кроме этого, группирование нескольких взаимосвязанных показателей в единую таблицу не всегда просто выполнить на основе исходных документов в формате HTML, Microsoft Word или Excel.

Проведенный на первом этапе работ комплекс предварительных исследований позволит заложить основу для создания и последующего развития и масштабирования ресурса. Предпринятые один раз, усилия коллектива УИС РОССИЯ обеспечат создание современного ресурса, способного поддержать новый уровень исследований по широкому кругу направлений и доступного всему научному сообществу России. Базу ресурса составят следующие источники:

- выборки из разделов сборников Госкомстата России «Российский статистический ежегодник», «Регионы России» и других;

- база данных «Паспорта городов»;
- данные Всероссийской переписи населения 2002 года;
- данные Министерства экономического развития и торговли РФ;
- данные Министерства финансов РФ;
- Атлас образовательных учреждений Министерства образования РФ.

Большинство коллекций содержит данные по 2300–2500 муниципальным образованиям в ретроспективе с 1996 года.

Кроме того, в состав ресурса будут включены коллекции, созданные исследовательскими коллективами и отдельными исследователями, к примеру, электронная версия книги «Город и деревня в Европейской России. 100 лет перемен», предоставленной Издательством «ОГИ» и авторами. Книга содержит большое число таблиц, причем данные проверены авторами. Ведутся переговоры с издательством «Новое издательство» о получении электронной версии книги «Сельская Россия на перепутье», разрешение автора Т.Г. Нефедовой получено. Книга также содержит статистические таблицы.

В 2004 году начата разработка первой версии реляционной базы на основе информационного фонда «Паспорта городов». Форма "Паспорта социально-экономического положения городов Российской Федерации" утверждена 1 июня 1999 г. В паспорт включено большое число показателей, сгруппированных по 23 разделам (население, труд и занятость, состояние окружающей среды, финансы и др.). Общий объем базы, включая сводный раздел, составляет 160 Мбайт. Коллектив проекта составляет данными по 1051 городу, данные включают около 50 показателей. Пользовательский интерфейс позволяет получить значение одного или нескольких показателей за выбранные годы с необходимой территориальной привязкой. Данные систематизированы с помощью единого каталога статистических показателей. Состав индикаторов, отображаемых в сводной таблице, периоды времени и перечень муниципальных образований выбираются пользователем. В настоящее время ведется разработка конструктора таблиц – дополнительного интерфейсного модуля, позволяющего формировать структуру таблиц в соответствии с потребностями исследователя. Конструктор позволит определять состав и иерархию полей, отображаемых по строкам и столбцам таблицы.

Одновременно начаты работы по экспертизе Классификатора статистических показателей (предоставлен Госкомстатом России, разработан с учетом международных стандартов). КСП включает около 5000 терминов, сгруппированных в 200 рубрик. Предполагается, что Классификатор станет дополнительным поисковым средством для базы данных по статистике. На следующем этапе коллектив УИС РОССИЯ предполагает интегрировать данные других статистико-

производящих государственных ведомств. Всего в РФ 37 ведомств, одно из направлений деятельности которых - сбор статистических данных.

3. База данных по муниципальной бюджетной статистике

Параллельно с реляционной базой по социально-экономической статистике в рамках УИС РОССИЯ создается реляционная база по бюджетной статистике в разрезе субъектов РФ и муниципалитетов. Реляционная база включает в себя данные по всем субъектам РФ и более чем 2 300 муниципальным образованиям (районы и города) за 1997–2002 годы. База содержит более 1 000 иерархических показателей, которые, по сути, являются группами, подгруппами, разделами, подразделами различных видов бюджетной классификации (классификации доходов, функциональной классификации расходов, классификации источников финансирования дефицита бюджетов). Дерево показателей представлено максимально подробно, что позволяет получать данные с глубиной до целевой статьи расходов или конкретного вида налогов. Например, можно не только получить данные о том, сколько средств поступало в тот или иной региональный бюджет от налога на прибыль предприятий и организаций, но и то, сколько средств в сумме этой доходной статьи составлял налог, к примеру, на доходы видеосалонов от проката видео и аудиокассет и записи на них. Если говорить о расходах, то база позволяет не только узнать, сколько средств тратилось из бюджета того или иного субъекта РФ или муниципалитета, например, на образование, но и сколько из них было потрачено на финансирование дошкольного, общего, профессионального, высшего и т.д. образования. С помощью базы данных можно сопоставлять плановые и исполненные данные по всем возможным видам бюджетов – консолидированному бюджету субъекта РФ, бюджету субъекта РФ, своду бюджетов муниципальных образований на территории субъекта РФ и бюджету конкретного муниципального образования, что с некоторыми оговорками (плановые задания могут меняться в течение года), позволяет сделать вывод о качестве краткосрочного бюджетного планирования в регионе.

В ходе работы над базой выполняется значительный объем исследований и предварительных процедур. Приведем только один пример. Данные по исполнению региональных и муниципальных бюджетов были выгружены из базы Министерства финансов РФ. Обработка и укладка их в реляционную базу потребовала построения иерархических деревьев показателей: деревья показателей расходов (функциональная классификация, ведомственная классификация, экономическая классификация), доходов,

источников финансирования дефицита, кредиторской задолженности. При этом, за период с 1997 по 2002 гг. дерево функциональной классификации расходов содержит около 5000 показателей; дерево экономической классификации расходов - около 1500 показателей; дерево ведомственной классификации расходов - около 6000 показателей (по 2000 г.); дерево доходов – около 300 показателей; дерево источников финансирования дефицита – около 50 показателей; дерево кредиторской задолженности – около 40 показателей. Идентификация иерархического уровня показателя и построение деревьев показателей производится по коду бюджетной классификации. Построение иерархических деревьев показателей происходило следующим образом. На первом этапе были построены деревья показателей за каждый год (с 1997 по 2002 г. – по 6 деревьев на каждый вид бюджетной классификации и кредиторская задолженность). В связи с тем, что из года в год бюджетная классификация претерпевала изменения, деревья показателей отличались друг от друга, причем иногда весьма существенно (менялись сами показатели – некоторые добавлялись, некоторые убирались, некоторые переименовывались). Самые серьезные изменения произошли в деревьях показателей по сравнению с 1997 годом.

На втором этапе происходило сравнение названий показателей, которым присвоен один и тот же код бюджетной классификации за год. При сравнении исследовалось несколько типичных ситуаций:

1. название показателя отличается несущественно. Например, 3000000000000110600: 1997 г. – оплата услуг по связи, 1998 и далее – оплата услуг связи.
2. название показателя отличается существенно. Например, 0904000363075000000: 2001 г. – текущее содержание подведомственных структур; 2002 г. – содержание подведомственных структур.
3. название показателя отличается существенно, причем одному и тому же названию (или похожему) в разные годы соответствуют разные коды классификации. Например, 010100000000000000 – в 1997 г. - Функционирование системы законодательной государственной власти; в 1998-2000 гг. - Функционирование главы государства - Президента Российской Федерации; в 2001 г. - Функционирование главы государства - Президента Российской Федерации (Функционирование президента республики в составе Российской Федерации и главы администрации субъекта Российской Федерации). В то же время в 1998-2002 гг. существовал показатель 010200000000000000 - Функционирование законодательных (представительных) органов государственной власти.

В первой ситуации исправлялось название показателя. Оно заменялось либо на более позднее, либо на более корректное с точки зрения грамматики. Среди показателей расходов таких случаев встретилось около 120. Во второй ситуации название показателя заменялось на более позднее, и к каждому показателю писалась справка о содержании показателя в прошлые годы. Среди показателей расходов таких случаев было около 370. Наконец, в ряде случаев 2а были написаны комментарии по загрузке (каким образом грузить тот или иной показатель).

Наконец, на третьем этапе по каждой из классификаций строилось единое дерево показателей, включающее в себя показатели за все годы. В каждом дереве проводилась корректировка иерархии показателей. Кроме того, при построении дерева показатели сопровождалась справками, содержащими информацию о том, за какие годы они присутствовали и как назывались ранее.

Кроме исследовательской, проводятся и другие виды работ, нацеленные на поддержку сервисов, помогающих в работе с бюджетной статистикой. К примеру, все сводные (построенные пользователем) таблицы могут быть импортированы в формате MS Excel. И, кроме того, непосредственно в режиме «on-line» возможен расчет важнейших вторичных показателей – на душу населения, удельный вес в валовом региональном продукте, удельный вес в показателях верхнего уровня (в процентах от общего объема доходов, расходов, дефицита и т.д.).

База создается как предметно-ориентированное направление УИС РОССИЯ и выполнена на основе единых технологических решений, что обеспечивает интеграцию бюджетных данных с 60 другими коллекциями системы, в том числе, социально-экономическими данными, нормативными документами федерального и регионального уровней, аналитическими докладами министерств и независимых исследовательских центров, публикациями в научных журналах и в СМИ.

Ведутся работы по созданию специального Рубрикатора по тематике «Бюджетная система РФ» как дополнительного поискового инструмента по интегрированной базе бюджетных данных и документам полного текста, в том числе, законы о бюджете (федеральные и субъектов РФ), отчеты о выполнении бюджетов, стенограммам Государственной Думы (пленарные заседания по обсуждению бюджета), аналитическим докладам, научным публикациям. Первая версия Рубрикатора уже реализована и используется для поиска по разделу «Статьи СМИ». К тестированию Рубрикатора привлечены специалисты Министерства финансов РФ, Института экономики РАН, Счетной палаты, Центра фискальной политики, нескольких муниципальных администраций.

4. Перспективы технологического развития

Следующий этап технологического развития УИС РОССИЯ – интеграция бюджетной и социально-экономической региональной и муниципальной статистики и создание сводного ресурса. Такой ресурс поддержит построение широкого круга вторичных показателей, на следующем этапе – автоматизированные сценарии анализа показателей, прогнозные модели оценки потенциала социально-экономического развития регионов на перспективу.

В рамках проекта будут разработаны новые методы анализа больших массивов данных социально-экономической и бюджетной статистики на основе специализированного программно-аналитического комплекса. Комплекс программных средств должен решать следующие задачи:

1. Поддержание интегрированной базы статистических данных. Структура базы будет построена таким образом, чтобы обеспечить соответствие типов представленных в ней данных применяемым методам их анализа.
2. Расчет вторичных переменных по стандартным алгоритмам и пользовательским формулам. В стандартный набор функций предполагается включить: показатели вариационного ряда (среднее значение, структурные средние, показатели вариации и концентрации признака и др.), рядов динамики (темпы роста и прироста, показатели сезонности и др.), статистическую оценку взаимосвязей (параметрические и непараметрические показатели тесноты связи) и ряд других индикаторов. Каждый показатель будет снабжен необходимыми методологическими комментариями, описывающими способы его расчета и интерпретации полученных результатов [2].
3. Визуализация данных, в том числе с использованием элементов геоинформационных технологий. Каждому типу исходных данных предполагается сопоставить набор рекомендуемых способов визуального анализа.
4. Поддержка сценариев анализа различных типов статистической информации. Для каждого типа статистических данных будет разработан сценарий анализа, включающий: 1) перечень рекомендуемых для расчета вторичных переменных; 2) набор рекомендуемых для построения сводных таблиц; 3) рекомендуемые способы визуализации данных. Каждый сценарий будет представлен в нескольких вариантах – типичный, краткий (с минимальным набором выходной информации) и развернутый (включающий максимальный рекомендуемый перечень показателей). Пользователь сможет редактировать сценарий исходя из особенностей решаемых задач.

5. Критериальный отбор муниципальных образований для проведения анализа. В качестве критерия может выступать соотношение реального и критического значения определенного показателя, выбранного пользователем, например: наличие дефицита (профицита) местного бюджета; удельный вес населения, имеющего денежные доходы ниже величины прожиточного минимума, и др. Пользователь сможет одновременно использовать несколько критериев отбора.

Функциональность системы будет определяться реальными потребностями специалистов. При этом будет использован практический опыт, накопленный в ходе чтения учебных курсов «Теория статистики» и «Экономическая статистика» на Экономическом факультете МГУ им. М.В. Ломоносова.

5. Социальная значимость проекта

Сводная база данных по социально-экономической и бюджетной статистике в разрезе страны – федеральных округов – регионов – муниципалитетов станет частью национальной информационной инфраструктуры России. Содержание и функциональные возможности обеспечат мониторинг ситуации в муниципалитетах, сопоставление показателей и сравнение динамики социальных, экономических, демографических процессов в каждом районе – городе – поселке и в нескольких районах одного региона или близлежащих регионов, что позволит выявлять закономерности и специфику ситуации и развития каждого образования, взаимозависимость тенденций, общие болевые точки. База данных будет иметь важную социальную значимость как первый источник, объединяющий данные на федеральном, региональном и местном уровнях. Ресурс будет представлен органам власти всех уровней, а также институтам местного самоуправления для использования при подготовке управленческих решений. С января 2006 года вступает в силу закон, по которому органы местного самоуправления получают полномочиями по принятию стратегических решений и наделяются правом самостоятельно формировать политику в вопросах местного значения. Для принятия обоснованных управленческих решений необходима полная и объективная информация о состоянии и динамике местной экономики и социальной сферы, проблемах и ресурсах для их разрешения. Эти же данные необходимы для поддержки общественного контроля и участия граждан в реализации муниципальных программ, именно активное и грамотное участие граждан в значительной мере определит действенность местного самоуправления. Доступ к надежной информационной базе способен оказать

существенную помощь органам местного самоуправления на этапе их становления в РФ.

На основе ресурса будут возможны исследования периферийных районов России и сельских районов. Интеграция с другими 60 коллекциями УИС РОССИЯ и подключение документальных источников расширяет исследовательский потенциал ресурса и его общественную значимость.

Литература:

- [1] Ветров Г.Ю., Визгалов Д.В., Шанин А.А., Шевырова Н.И. Индикаторы социально-экономического развития муниципальных образований. Издание 2-е, дополненное. // Фонд «Институт экономики города». – М., 2002.
- [2] Теория статистики. / Под ред. Г.Л. Громыко. – М.: ИНФРА-М, 2000.
- [3] Юдина Т.Н. и др. Предметно-ориентированные ресурсы Университетской информационной системы РОССИЯ – базы данных «Статистика России» и «Бюджетная система Российской Федерации». / Труды Всероссийской научной конференции «Научный сервис в сети Интернет». – М.: Изд-во МГУ, 2003.

University Information System RUSSIA: Subject-oriented Data Bases to Integrate Social, Economic, Budget Data at Federal, Regional, Local Levels

© Anna Bogomolova^{2,3}, Oleg Karasev^{2,3},
Ruslan Sennov^{2,4}, Tatyana Yudina^{1,2}

1) Moscow State University
Research Computing Center

2) Non-profit Center for Information Research

3) Moscow State University Economic Faculty

4) Moscow State University Physics Faculty
yudina@mail.cir.ru

The article describes two databases that are currently under construction as University Information System RUSSIA (www.cir.ru) - subject-oriented holdings "Statistics of Russia" and "RF Budget system" – the databases' concepts, investigations, design, stages of implementation, content and value-added service to compliment the data and documents. Both databases integrate local level statistics. The main accomplishment is relational database format for socio-economic and budget data available at federal, regional and local levels. Each indicator may be monitored at 10 years coverage and analyzed exploiting applied math methods and developed models. System and comparative investigations are possible providing for social analyses in support for decision making at federal, regional, local level.

* Работа поддержана грантом РГНФ № 04-02-12004в