

# Университетская информационная система РОССИЯ: социально-экономическая и бюджетная статистика в формате реляционной базы, сервисы для системного и сравнительного анализа\*

Богомолова А.В.<sup>2,3)</sup>, Карасев О.И.<sup>2,3)</sup>, Сеннов Р.А.<sup>2,4)</sup>, Юдина Т.Н.<sup>1,2)</sup>

<sup>1)</sup> Научно-исследовательский вычислительный центр МГУ им. М.В. Ломоносова

<sup>2)</sup> АНО Центр информационных исследований

<sup>3)</sup> Экономический факультет МГУ им. М.В. Ломоносова

<sup>4)</sup> Физический факультет МГУ им. М.В. Ломоносова

## Аннотация

Статья описывает новый продукт, разработанный в рамках Университетской информационной системы РОССИЯ (uisrussia.msu.ru) - реляционную базу, в которой интегрирована государственная социально-экономическая статистика Росстата и бюджетная статистика Министерства финансов и Федерального Казначейства. Ресурс поддерживает системный и сравнительный анализ на уровне регионов и муниципалитетов, включает развитые средства визуализации данных.

## 1. Статистическая система нового поколения как компонент национальной информационной инфраструктуры

В последние годы страны-лидеры приняли программы развития национальных информационных инфраструктур 21 века. Одно из основных направлений – разработка статистических систем нового поколения, в которых интегрированы ресурсы органов государственной власти всех уровней, и доступны сервисы, обеспечивающие системный и сравнительный анализ на всех уровнях управления страной. В дополнение к статистической базе, способной поддерживать системный анализ социальных, демографических, экономических, бюджетных, экологических процессов на основе комплексного исследования всего состава показателей государственной статистики (data intensive research), разрабатываются технологии и лингвистические ресурсы (рубрикаторы и онтологии) для содержательной интеграции статистических показателей и тематических аналитических публикаций (knowledge products) государственных ведомств, университетов и исследовательских центров. Разработка статистических систем нового поколения ведется

государственными организациями в сотрудничестве с университетами, координируют работы национальные научные фонды. Национальный научный фонд США в 2005 году объявил специальную программу Cyberinfrastructure, по которой ведутся фундаментальные и прикладные исследования в области информатики, ориентированные на формирование в стране общества знаний, разрабатываются новые методики и технологии системного социального анализа. [1]

Ряд проектов по созданию информационных ресурсов нового поколения в США, Канаде, Великобритании, Финляндии, получивших государственную поддержку через бюджет национальных программ, были инициированы и велись университетами в рамках создания научной информационной инфраструктуры для учебных и исследовательских задач. Фактически университеты ранее уже создали прототипы информационных систем и аналитических сервисов, которые стали основой новой информационной инфраструктуры своих стран.

Как свидетельствует опыт ряда стран, по мере роста общей информационной культуры в стране, повышения уровня компьютерной образованности государственных служащих и общего уровня государственного управления, возрастает потребность в качественных информационных ресурсах, дополненных развитыми технологиями для системного социального анализа. На этом этапе управленческие структуры всех уровней проявляют растущую заинтересованность в сотрудничестве с университетами в целях использования новых методик социального анализа и прогноза при разработке государственных программ и конкретных решений.

Как и в других странах, в университетах и научных организациях России создаются информационно-аналитические ресурсы для полноценных социальных исследований на основе обработки больших массивов данных,

разрабатываются методы комплексного анализа социальных, демографических, экономических, бюджетных, экологических процессов.

Пока востребованность ресурсов и знаний научного сообщества России органами власти для задач государственного управления невысока. Остается надеяться на новое поколение управленцев и на новые образовательные программы в университетах и вузах, которые включают как обязательные предметы обучение методам системного социального анализа с использованием баз данных для задач государственного управления.

## 2. Реляционная база для анализа государственной социально-экономической и бюджетной статистики РФ

Один из проектов по созданию ресурсов и сервисов для системного социального анализа государственной статистики выполняется в рамках Университетской информационной системы РОССИЯ (УИС РОССИЯ, uisrussia.msu.ru/www.cir.ru). В 2006 году в составе УИС РОССИЯ реализована реляционная база, интегрирующая государственную социально-экономическую статистику Росстата и бюджетную статистику Министерства финансов и Федерального Казначейства. Данные доступны в разрезе субъектов федерации и муниципалитетов. Реляционная база содержит 15 тыс. показателей, общий объем 1,7 Гб

Интерфейс реляционной базы предполагает пошаговый выбор:

- 1) территории
- 2) года
- 3) вида бюджета
- 4) показателей бюджетной и социально-экономической статистики.



Рис. 1. Пошаговое построение запроса

В реляционной базе представлены данные по бюджетной статистике в разрезе субъектов РФ и муниципальных образований в ретроспективе с 1998 г. по 2005 г. Доступны данные об исполнении бюджетов более чем 2000 муниципалитетов.

В части бюджетной статистики база включает в себя как плановые данные, так и данные об исполнении бюджетов в разрезе консолидированного бюджета субъекта РФ, бюджета субъекта РФ, бюджета муниципального образования.

Показатели сгруппированы в четыре раздела – показатели доходов бюджетов; показатели расходов бюджетов; показатели дефицита бюджетов и источников его финансирования; показатели социально-экономической статистики. Группировка доходов, расходов, дефицита бюджета и источников его финансирования основана на бюджетной классификации, применявшейся в период с 1998 по 2004 гг.

В соответствии с Бюджетным кодексом РФ бюджетная классификация представляет собой группировку доходов, расходов, источников финансирования дефицитов бюджетов всех уровней бюджетной системы РФ, а также видов государственного (муниципального) долга и государственных (муниципальных) активов [2]. Бюджетная классификация используется для составления и исполнения бюджетов всех уровней бюджетной системы РФ и должна обеспечивать сопоставимость показателей бюджетов всех уровней бюджетной системы РФ за различные годы. Однако на протяжении всего рассматриваемого периода в бюджетную классификацию ежегодно вносились существенные изменения, в результате чего на этапе подготовки показателей для загрузки в реляционную базу каждый год проводится экспертная оценка произошедших в бюджетной классификации изменений и их анализ. С 2005 г. бюджетная классификация претерпела столь существенные преобразования, что в настоящее время поддерживается две различные базы с различными интерфейсами – для данных с 1998 по 2004 гг. и для данных с 2005 г.

В процессе разработки базы данных существенной технической проблемой была и остается разноформатность исходных данных, получаемых из субъектов РФ. Различные субъекты РФ используют разные форматы выгрузки данных – txt, xls, xml, dbf. При этом, например, в файлах формата xls встречается 4 варианта таблиц. Таким образом, на подготовительном этапе проводится как анализ исходных данных, так и преобразование всех данных в текстовый формат с заданным количеством полей.

Источником показателей социально-экономического развития субъектов РФ является ежегодный сборник, публикуемый Федеральной службой государственной статистики (Росстатом), – «Регионы России» за 1998-2005 гг. Показатели организованы в иерархическое дерево на основе разделов и подразделов сборников Росстата.

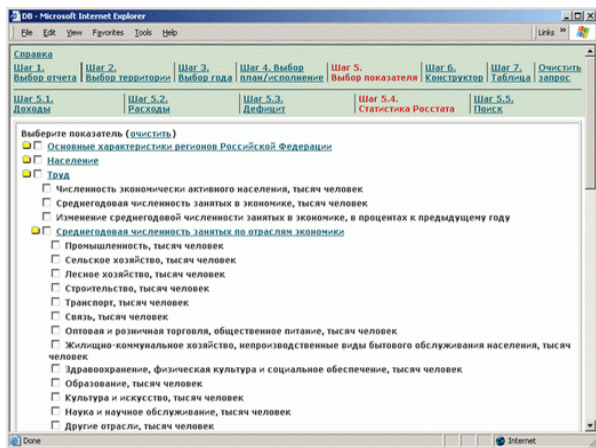


Рис. 2. Выбор показателей социально-экономического развития субъектов РФ

База содержит ряд сервисных возможностей для поддержки работы с таблицами. Одной из таких возможностей является конструктор таблиц, позволяющий задавать содержание строк и столбцов таблицы, а также выбор вспомогательных кодов.

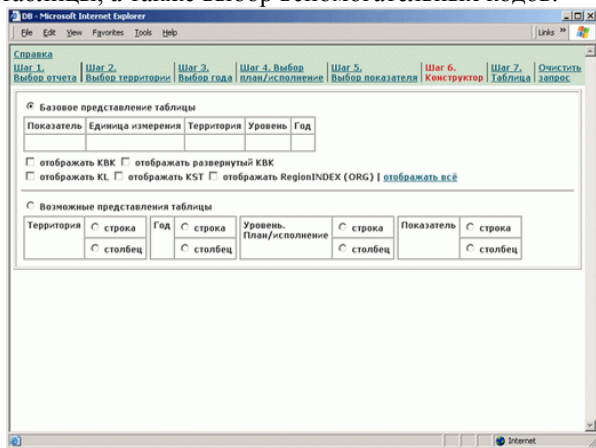


Рис. 3. Конструктор таблиц

Все таблицы могут быть получены в формате MS Excel.

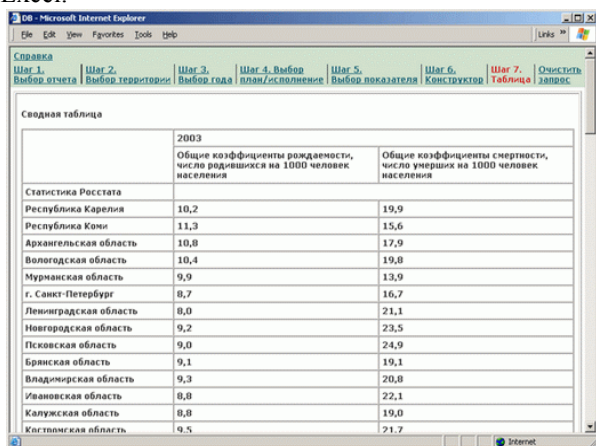


Рис. 4. Отображение данных в таблице

Для всех таблиц, отображающих данные в разрезе субъектов РФ, может быть построена карта-схема.

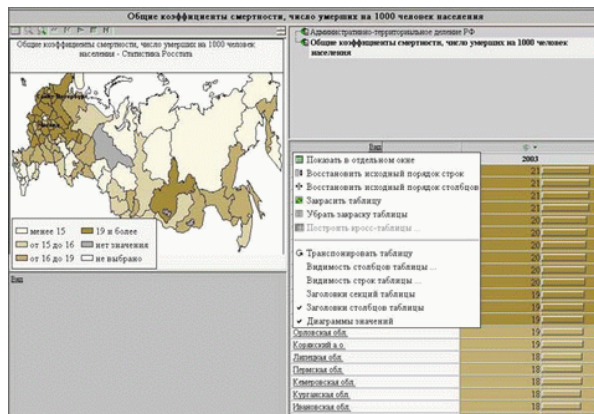


Рис. 5. Отображение данных на карте

Карта-схема позволяет не только визуально представлять данные, но и работать с таблицей, осуществляя сортировки, скрывая или отображая строки и столбцы таблицы. Также предусмотрена возможность транспонировать таблицу, строить диаграммы значений, изменять группировку данных.

Интерфейс реляционной базы оптимизирован для просмотра в браузере Internet Explorer 5.5 и выше. На данный момент в качестве сервера используется PC Pentium II (частота – 700 МГц, оперативная память – 512 Мб) под управлением операционной системы ALT-Linux 2.4.26, со стандартными серверными компонентами: WEB-сервер Apache 1.3.33, база данных MySQL 4.0.24, язык программирования PHP 4.3.10.

Разработан специальный интерфейс реляционной базы для открытого доступа на сайте «Бюджетная система РФ» <http://www.budgetrf.ru/openDB/index.php>

Открытая часть реляционной базы содержит около 600 показателей бюджетной и 100 показателей социально-экономической статистики.

### 3. Перспективы проекта

Предполагается, что в 2006-2007 годах в реляционную базу будет добавлена статистика, собираемая другими государственными ведомствами РФ и статистика, предоставляемая региональными органами власти.

Следующий этап работ по проекту – интеграция тематических публикаций – докладов государственных органов и исследовательских центров, научных изданий, материалов СМИ. Будут включены также публикации на английском языке – доклады международных организаций, научные статьи. Для содержательной интеграции статистических показателей и публикаций создан рубрикатор и разрабатывается онтология «государственные финансы РФ».

В 2005-2006 годах в работе по проекту участвовал студент факультета Вычислительной Математики и Кибернетики МГУ им. М.В. Ломоносова Кудинов П.

## Литература

- [1] F.Berman and H.Brady. National Science Foundation. Workshop on Cyberinfrastructure for Social and Behavioral Sciences. Final Report, 2005 (<http://www.sdsc.edu.sbe/>)
- [2] Бюджетный кодекс Российской Федерации от 31 июля 1998 г. N 145-ФЗ // <http://www.garant.ru>

**University Information System RUSSIA:  
Relational Database Integrating RF  
Economic, Social and Budget Statistics.  
Value-added Services for System and  
Comparative Investigations**

Bogomolova A.V., Karasev O.I., Sennov R.A., Yudina T.N.

Lomonosov Moscow State University Research  
Computing Center

Non-commercial Center for Information Research  
Lomonosov Moscow State University Economic  
Faculty

Lomonosov Moscow State University Physics Faculty  
Article describes a new product implemented under the  
University Information System RUSSIA  
([uisrussia.msu.ru](http://uisrussia.msu.ru)) - relational database that integrates  
RF economic, social and budget data obtained from  
Rosstat, Ministry of Finance and Federal Value-added  
services are available. In 2006-2007 other RF state  
agencies data will be added.

---

\* В 2004-2006 годах работа поддерживается грантом  
РГНФ № 04-02-12004в, грантом РГНФ 05-02-12206в,  
грантом РФФИ № 04 07-90279/в