

Абылкаликов С. И., Москва

Сравнительный анализ основных источников данных о миграции населения России

Аннотация

В статье проводится сравнительный анализ источников данных о миграции в России, публикуемых Росстатом и иностранными статистическими службами. Основное внимание уделяется сопоставимости данных и удобству работы с ними. Выявленные недостатки будут минимизированы в разрабатываемой Институтом демографии НИУ ВШЭ База данных демографических показателей по регионам России и странам мира.

Ключевые слова: миграции населения, Росстат, база данных, демографические показатели

В НИУ ВШЭ с 2011 года стартовал исследовательский проект «Разработка базы данных демографических показателей по регионам России и странам мира» №11-04-0039 по конкурсу Программы «Научный фонд ВШЭ» «Учитель – Ученики 2011-2012». Целью Проекта является объединение основных источников демографических показателей и предоставление доступа к накопленной информации в едином удобном формате через гибкий веб-интерфейс. Важнейшим этапом данной работы является анализ имеющихся источников данных, выявление их особенностей и недостатков, анализ интерфейса пользователя, оценка источников по широкому спектру критериев (характеру доступа, точности, форматам, периодичности обновления и другим).

Рассмотрим три основных источника официальных статистических данных по миграции, выпускаемых Федеральной службой государственной статистики Российской Федерации (Росстатом): статистический сборник «Демографический ежегодник России» (Демежегодник), бюллетень «Численность и миграция населения Российской Федерации», а также электронную «Центральную базу статистических данных» (ЦБСД). Первые два источника имеют бумажную и электронную версии, выпускаемые ежегодно, ЦБСД имеет только электронную версию.

Данные в ЦБСД объединены в 16 блоков, от государственных финансов и национальных счетов, до статистики по труду и занятости. В блок «Демография» входят 10 разделов, включая раздел миграции насе-

ления. Показатели по миграции можно объединить в три группы: прибытие, выбытие и миграционный прирост, которые можно рассмотреть по полу, возрасту, направлениям (регионам, странам) прибытия или выбытия и коэффициенты миграции (на 10 тыс. чел.). Также в ЦБСД содержатся данные по миграции в районах Крайнего Севера и приравненных к ним территорий и в местах радиоактивного загрязнения в результате аварии на Чернобыльской АЭС. В дежегоднике содержатся статистические данные об изменении численности и возрастно-половом составе населения, его размещении на территории России, о рождаемости и смертности, брачности и разводимости, а также миграции. В дежегоднике имеются показатели прибытия, выбытия и миграционного прироста с характеристиками, которые в том числе отсутствуют в ЦБСД. К примеру, возрастно-половой состав, гражданство, национальность международных мигрантов и т. д. Бюллетень содержит сведения о численности населения и общих итогах миграции по России и ее субъектам, по районам Крайнего Севера, зонам радиоактивного загрязнения, вынужденных переселенцах и беженцах, данные о численности участников Государственной программы содействия добровольному переселению в Российскую Федерацию соотечественников, проживающих за рубежом (с 2007 года), а также данные о миграции по целям поездок, по обстоятельствам переезд, по видам миграции и длительности проживания в предыдущем месте жительства. [1].

В демографическом ежегоднике можно встретить данные, начиная с 1897 года (численность и состав населения), однако, данные по миграции доступны лишь с 1993 года. В каждом выпуске содержатся данные по нескольким годам, но для составления непрерывного ряда придется изучить сразу несколько выпусков. В электронном виде на сайте Росстата доступны выпуски, начиная с издания 2001 года. Данные в ЦБСД по большинству показателей размещены с 1990 года по настоящее время за каждый год. Кроме уже указанного временного интервала, в некоторых случаях, данные представлены с 1970 по 1990 годы с пятилетним интервалом, к примеру, такой показатель, как числа родившихся за год. Бюллетень «Численность и миграция населения Российской Федерации» издается с 1993 года, данные по некоторым показателям сравниваются с показателями предыдущего года. В связи с тем, что на сайте Росстата присутствуют электронные версии лишь 2009, 2010 и 2011 годов выпуска, временной горизонт ограничен 2007-2010 годами, причем для составления непрерывного ряда данных придется воспользоваться всеми имеющимися выпусками.

В Дежегоднике данные представлены в html (выдача таблицы на экране в браузере) и файлах формата doc (Word). Файлы формата doc сгруппированы по главам и собраны в один zip-архив. Выпуск 2010 года целиком представлен в формате pdf. Бюллетень также представлен в форматах html и doc, также имеется возможность получения xls (Excel) файлов. Файлы doc и xls организованы по главам, которые собраны в zip-архив. При запросах в Центральную базу статистических данных можно получить данные в следующих форматах: html (веб-страницы), csv (текстовый файл с разделителем) и xls. Большие массивы данных для дальнейшей работы с ними удобнее

всего использовать в форматах xls и csv. Неудобным являются копирование с html - в некоторых интернет-браузерах при копировании нарушается табличная структура (например, в Mozilla Firefox, Opera и Google Chrome), файлы формата doc данных публикаций, как правило, содержат большое множество лишних пробелов, знаков абзаца, прочих незначащих знаков, а также ручных переносов. Бумажные версии удобны для тех, кому привычнее работать с книгами, чем с электронными носителями информации и для просмотра нескольких отдельных показателей.

ЦБСД, как и любая онлайн-база данных, может обновляться часто и без ограничений. Однако, в отличие от других баз, в частности Базы данных Евростата [2], в ЦБСД отсутствуют указания даты последнего обновления. Демежегодник и бюллетень «Численность и миграция населения Российской Федерации», несмотря на наличие электронных версий, являются ежегодными бумажными изданиями, поэтому информация в них обновляется лишь с выходом нового издания. При этом бывает так, что статистическая информация, данная в выпусках разных годов, не совпадает между собой, т. к. перерасчет данных, уточненных по итогам переписей населения производится по большей части за счет миграции [6]. Так, миграционный прирост всего населения России в 2000 году составил 217 975 чел. по данным демежегодника-2001 и 241 755 чел. по данным демежегодника-2002 и всех последующих [3, 4].

Таким образом, даже в источниках Росстата демографические показатели могут иметь разные значения. Если взять тот же самый показатель миграционного прироста в России, опубликованных в иностранных источниках, таких как INED Developed countries database и базе данных Евростата (см. таблицу 1), то можно выяснить, что значение показателя для одного и того же периода не всегда является однозначным. Различия показателей в разных источниках могут иметь разные причины, связанные как с методикой подсчета, округлением до разных рангов и т. д. Однако, выяснить, какие источники содержат «правильные» показатели, а какие «ошибочные» в случае несовпадения данных, по-видимому, невозможно.

Таблица 1

Значение показателя миграционный прирост в России по данным различных источников, чел.

	2004	2005	2006	2007	2008	2009
ЦБСД Росстата	41275	107432	132319	239943	242106	247449
INED Developed countries database	98939	125891	154483	258193	257148	259956
База данных Евростата	98939	125890	132319	239956	242106	247449

1 – ЦБСД Росстата [7], 2 – INED Developed countries database [11], 4 – База данных Евростата [2] доступ 12.04.2012.

Разрабатываемая База данных должна обладать следующими свойствами:

- Множественностью значений показателей, полученных из различных источников. Пользователи должны иметь возможность формировать результат запроса к Базе данных с уточнениями, в том случае, если База данных содержит множественные значения данного показателя, но без дублирования одинаковых данных (к примеру, подготовленных одним ведомством для одного периода).

- Широким перечнем показателей. К примеру, деможегодник, ЦБСД и бюллетень «Численность и миграция населения Российской Федерации» содержат как одинаковые показатели, так и уникальные, содержащиеся только в них.

- Охватывающей как можно больший временной ряд. Если источник (к примеру, бюллетень «Численность и миграция населения Российской Федерации») имеет данные только на 1 год, должна быть возможность автоматического составления непрерывных рядов данных.

- Полнотой рядов данных. Если какие-то массивы цифр отсутствуют (ИНЕД с 1950 г., или бюллетень «Численность и миграция населения Российской Федерации»), они могут быть дополнены из других источников (например, данные в ИНЕД доступны начиная с 1950 года).

- Содержанием информации не только из других баз данных, но и уникальных данных из таких источников, как ежегодные демографические доклады «Население России» и еженедельный демографический Интернет-журнал «Демоскоп Weekly».

- Единым стандартом получения данных, с возможностью копирования. Формат xls является более предпочтительными, чем csv в силу того, что на современных машинах нехватка постоянной памяти уже не является непреодолимым препятствием.

- Работой без сбоев. База данных должна обладать надежной программной оболочкой и находиться на надежном сервере.

- Своевременным обновлением, с указанием даты последнего обновления.

- Возможностью получения прямой ссылки на массивы данных.

- Интуитивно понятным интерфейсом.

- Бесплатностью доступа.

ЦБСД является крупнейшим источником официальных статистических данных по Российской Федерации, который содержит самую актуальную информацию. База данных будет полезна тем, кому нужны общие показатели по всем направлениям статистики. Если требуется знать, как оценивались те или иные показатели в разное время, то лучше воспользоваться бумажными изданиями. В деможегоднике, в отличие от ЦБСД, содержатся специфические демографические показатели, включая и ми-

грацию населения. Для составления полного временного ряда придется воспользоваться несколькими изданиями. Бюллетень «Численность и миграция населения Российской Федерации» является специализированным изданием по статистике миграции. Он включает в себя как показатели, встречающиеся в других источниках, так и уникальные разработки, публикуемые только в этом издании. Существенным недостатком бюллетеня является сложность сравнения динамики показателей по годам, т. к. в каждом выпуске оцениваются лишь текущий и предыдущий годы.

Все три описанных источника являются бесплатными и находятся в открытом доступе. Однако, они имеют недостатки и особенности, затрудняющие работу с содержащейся в них информацией. Для преодоления этих недостатков разрабатывается База Данных Института демографии НИУ ВШЭ, которая будет содержать большое количество демографических данных, наиболее часто используемых в исследованиях и предоставлять неограниченный и бесплатный доступ в удобном формате через гибкий веб-интерфейс.

Библиографический список

1. База данных ИДЕМ ГУ-ВШЭ URL: http://db.demoscope.ru/bd_sources_di.php.
2. База данных Евростата URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.
3. Демографический ежегодник России 2001 // Статистический сборник. М.: Росстат. 2001.
4. Демографический ежегодник России 2002 // Статистический сборник. М.: Росстат. 2002.
5. Демографический ежегодник России 2010 // Статистический сборник. М.: Росстат. 2010.
6. Мкртчян Н.В. Статистические источники информации о миграции населения в России // Демоскоп Weekly. №335-336. 2008. URL: <http://www.demoscope.ru/weekly/2008/0335/analit02.php>.
7. Сороко Е.Л. База данных демографических показателей по регионам России и странам мира: принципы построения// Текст доклада на Четвертой Международной научной конференции «Инновационное развитие экономики России: институциональная среда». Экономический факультет МГУ имени М.В. Ломоносова. 20–22 апреля 2011 г.
8. Центральная база статистических данных Росстата URL: <http://www.gks.ru/wps/wcm/connect/rosstat/rosstatsite/main/database/cbsd/>.
9. Численность и миграция населения Российской Федерации в 2010 году // Статистический бюллетень. М.: Росстат. 2010.

10. Юмагузин В.В. Анализ источников демографической информации как важнейший этап разработки базы данных демографических показателей // Доклад на 2-й Международной научно-практической конференции студентов и аспирантов «Статистические методы анализа экономики и общества» М.: НИУ ВШЭ, 2011.
11. INED Developed countries database URL: http://www.ined.fr/en/pop_figures/developed_countries/developed_countries_database/.