БАЗЫ ДАННЫХ И ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ

DATABASES AND INFORMATION RETRIEVAL SYSTEMS

ИСТОРИКО–ДЕМОГРАФИЧЕСКИЕ БАЗЫ ДАННЫХ В КОНТЕКСТЕ ЕВРОПЕЙСКИХ КОМПАРАТИВНЫХ ИССЛЕДОВАНИЙ: ПРОЕКТ EHPS–NET*

HISTORICAL-DEMOGRAPHIC DATABASES IN THE CONTEXT OF EUROPEAN COMPARATIVE STUDIES: EHPS-NET PROJECT

Горбачев Олег Витальевич,
доктор исторических наук, профессор,
профессор кафедры документационного
и информационного обеспечения управления
департамента «Исторический факультет»
Института гуманитарных наук и искусств
Уральского федерального университета
им. первого Президента России Б. Н. Ельцина
E-mail: og_06@mail.ru

Oleg V. Gorbachev

В связи с проблемой координации усилий историков-демографов в статье рассказывается о проекте Европейской сети первичных исторических данных по народонаселению (EHPS-Net), начатом в 2011 г. Предложенная разработчиками структура промежуточных данных (IDS), не ограничивая возможности исходных национальных и региональных массивов данных, позволит использовать их потенциал в проведении межстрановых исследований истории семьи, рождаемости, смертности и других демографических процессов. Характеризуется структура EHPS-Net, основные направления ее деятельности, современное состояние проекта. Ставится вопрос о необходимости российского участия в программе, а также в работе вновь созданного Европейского общества исторической демографии.

The article considers the project of European Historical Population Samples Network (EHPS-Net) launched in 2011. Its developers have proposed intermediate data structure (IDS), which does not limit the possibilities of available national and regional data sets, but allows using their potential in cross-country studies on family history, fertility, mortality and other demographic processes. The structure of EHPS-Net, its main activities and the current state of the project are observed. The question of Russian scholars' participation in the program as well as in the work of the newly established European Society of Historical Demography (ESHD) is discussed.

* Публикация поддержана грантом Российского фонда фундаментальных исследований 15–06–05611 «Типология раннесоветской семьи по материалам партийных переписей 1920-х гг.».

Ключевые слова: история народонаселения, первичная демографическая информация, историко-демографические базы данных, сравнительные демографические исследования, международное научное сотрудничество.

Keywords: history of population, the primary demographic information, historical and demographic databases, comparative demographic studies, international scientific cooperation

Те, кто занимается созданием больших историко-статистических баз данных, хорошо знают, что одной из наиболее болезненных проблем является сопоставимость показателей, полученных в рамках разных проектов. Поскольку российские архивы дают возможность работы главным образом с агрегированными данными, а первичные сведения утрачены либо рассеяны, вопрос зачастую сводится к установлению методики получения вторичной информации с целью ее последующей унификации с уже имеющейся.

Проблема обработки первичных данных решается каждым исследователем индивидуально, но лишь до тех пор, пока не возникает вопрос об их интерпретации в более широком контексте.

В отличие от России, в архивах других стран обширные массивы доступных первичных данных не являются редкостью. Речь главным образом идет о демографической информации (переписи населения, специальные переписи, данные текущего учета и т. п.). В настоящее время предпринимаются активные усилия по объединению имеющихся материалов и обеспечению свободного исследовательского доступа к ним посредством сети Интернет в рамках концепции Big Data. Движущими силами этого процесса следует считать глобализацию и новые возможности информационной среды.

Наиболее масштабным проектом подобного рода является американский IPUMS (Integrated Public Use Microdata Series)[1]. Он реализуется Центром народонаселения при Университете штата Миннесота и опирается на первичные данные переписей населения, предоставляемые национальными статистическими управлениями. На сегодняшний день проект охватывает 76 стран, в том числе четыре страны постсоветского пространства (Армению, Беларусь, Киргизию и Украину).

При всех достоинствах глобальных баз демографических данных их издержки очевидны — унифицированный формуляр отсекает «лишнюю» информацию, несмотря на ее потенциальную исследовательскую ценность. Поэтому европейские историки-демографы пошли по пути «мягкой» унификации для сохранения вариативности доступных демографических сведений. С этой целью в июне 2011 г. под эгидой Европейского научного фонда (European Science Foundation, ESF) была создана исследовательская сеть, объединившая историков народонаселения, работающих с первичной информацией (European Historical Population Samples Network, EHPS-Net[2]). Программа проекта рассчитана на пять лет, до июля 2016 г. Ее руководителем является профессор Кис Мандемакерс (Kees Mandemakers) из Международного института социальной истории (Нидерланды).

EHPS-Net призвана объединить ученых с тем, чтобы создать общий формат баз данных, содержащих неагрегированную информацию об отдельных лицах, а также семьях и домохозяйствах. На выходе предполагается получить формат промежуточных данных (Intermediate Data Structure, IDS). Он будет играть роль интерфейса между разрозненными европейскими базами данных для возможности выхода за пределы регионального или национального уровня.

С целью обеспечения разных видов исследований (например, по вопросам миграции и рождаемости) в тесном сотрудничестве исследователей и программистов разрабатываются специальные алгоритмы извлечения данных. Основными принципами работы будущей IDS являются открытость, масштабируемость и расширяемость. Заявляется, что каждый желающий имеет возможность добавить свой модуль работы с информацией, который пройдет независимую экспертную оценку и будет представлен в издаваемом Сетью электронном журнале «Historical Life Course Studies»[3].

В настоящее время на сайте EHPS-Net создан портал, предоставляющий доступ к европейским базам данных, а также к некоторым неевропейским, также присоединившимся к проекту. Всего на начало 2015 г. представлено 22 БД, из них 7 неевропейских. В числе представленных — базы Кембриджской группы истории населения (Шотландия), Норвежского центра исторической информации при Университете Тромсё, проекта «Mosaic» (Институт Макса Планка, Германия), Международного института социальной истории (Нидерланды), Университетов Техаса и Юты (США), Квебека (Канада), базы «POPUM» Университета Умео (Швеция), база «Старая Карелия» (Karelia Database Foundation, Финляндия) и др. Самые ранние данные датируются 1430 г. («Mosaic»). Наиболее поздние сведения содержат экономико-демографическая база данных Университета Лунда (Швеция) (до 2011 г.) и база данных Университета штата Юта (США) (по настоящее время) (рис. 1)[4].

Рис. 1. Распределение демографических баз данных, присоединившихся к EHPS-Net, по странам Европы (февраль 2015 г.)

На портале дан обзор всех баз данных, где, среди прочего, оценивается степень их соответствия формату IDS. Здесь же предполагается выложить документацию по стандартизации данных и программное обеспечение, позволяющее конвертировать отдельные базы в формат IDS. Разработчики рассчитывают, что количество программ-конвертеров будет достаточно велико. Часть информации будет доступна только участникам проекта.

Координаторы проекта особо отмечают, что его целью является не замена существующих баз данных некоей «центральной» базой, а оптимизация использования уже имеющихся сведений.

В качестве одной из целей проекта заявлено стимулирование создания новых историко-демографических баз данных, поскольку их сегодняшнее количество явно недостаточно. Сеть позволит объединить усилия тех ученых, которые до настоящего момента действуют разрозненно, в рамках национальных историко-демографических традиций. Это будет способствовать углубленному пониманию истоков проблем современного европейского и мирового демографического развития.

Различия в доступных базах данных обусловлены разнообразием источников, содержащих персональные сведения. Это переписи населения, списки налогоплательщиков, данные церковного учета и т. п. Кроме того, каждый такой массив источников имеет свои особенности, связанные с историческими условиями, целями и возможностями документирования. Различия в содержании не позволяют исследователю получать одинаково содержательные ответы на свои вопросы к имеющимся базам данных. Проще всего, как показывает опыт IPUMS, в общий массив включаются данные переписей населения. Фактически они уже готовы для обработки в IDS.

Каким образом предполагается достичь сопоставимости данных? Стратегия работы с демографической информацией в рамках EHPS-Net представлена на рисунке 2[5].

В левой части схемы показаны источники информации (регистры населения, посемейные списки, сведения о налогоплательщиках и т. д.). На их основе составляются национальные базы данных за определенный период (второй столбик). На сле-

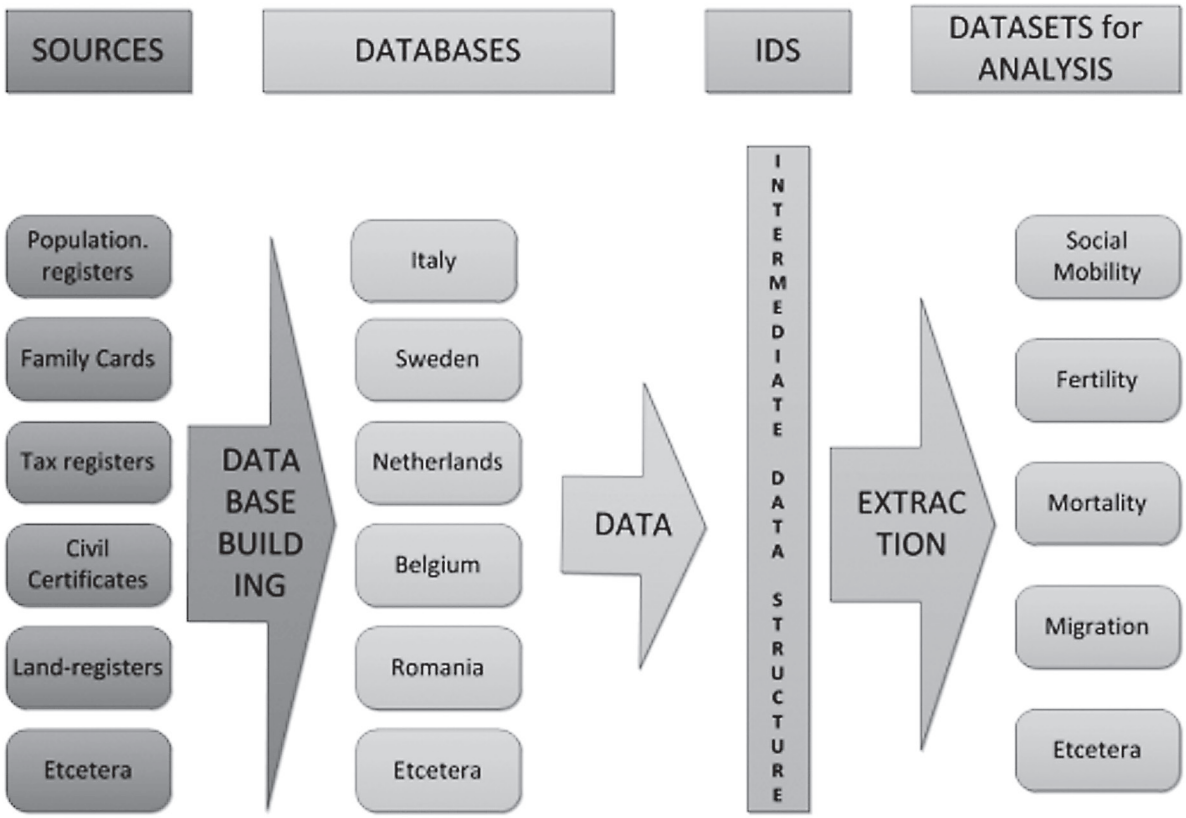

Рис. 2. Схема работы с данными в EHPS-Net

дующем этапе подключается EHPS-Net, предлагая свой формат IDS для этих данных. На выходе (правая часть схемы) — пригодные для исследовательского анализа массивы данных по социальной мобильности, рождаемости, смертности и т. п. Задача, таким образом, распадается на две части:

1) оздание программного интерфейса для конвертации национальных (региональных) баз данных в формат IDS;
2) создание собственно IDS с программными возможностями, позволяющими создавать тематические массивы данных[6].

Важным достоинством IDS является то, что она предполагает использование стандартных процедур для формирования тематических выборок по исследовательским запросам, что серьезно повышает эффективность научной работы. На заключительном этапе проекта предполагается выложить на его сайте «пилотные» результаты в виде совокупности данных для анализа миграции, рождаемости и социальной мобильности. Они должны помочь в точной настройке среды IDS.

В настоящее время в структуре EHPS-Net имеется восемь рабочих групп (до недавнего времени их было десять). В их обязанности входят разработка сайта (группа 1), издание электронного журнала (группа 2), разработка программного обеспечения для IDS (объединенные группы 3, 4 и 10), работа с государственными и общественными структурами, фондами, обеспечение финансовой поддержки (группа 5), привлечение новых баз данных (группа 6), обеспечение образовательных программ (группа 7), разработка стандартов документации по ведению баз данных (группа 8), работа с ГИС (группа 9).

Ежегодно проводятся общие собрания EHPS-Net. В третьем и пока последнем собрании, прошедшем в сентябре 2014 г. в Альгеро (Италия), приняли участие исследователи из 16 стран, в том числе впервые из России. Ее представляли члены АИК из Уральского федерального университета Е. М. Главацкая, О. В. Горбачев и Л. Н. Мазур. Российскими участниками был заявлен проект создания нескольких баз данных в рамках недавно созданного на историческом факультете Международного центра демографических исследований (МЦДИ). Руководителем центра является активный деятель EHPS-Net профессор Гюннар Торвальдсен, возглавляющий также Норвежский центр исторической информации при Университете Тромсё.

Наиболее очевидные возможности создания баз данных, совместимых с IDS, дают доступные на Урале церковные метрические книги конца XIX — начала XX в., а также «ревизские сказки» XVIII–XIX вв. Кроме того, в распоряжении россий-

ских исследователей имеются некоторые первичные материалы советского времени, в том числе данные Всесоюзной переписи населения 1959 г. Реализация этих планов позволит соотнести российские демографические процессы с европейскими и мировыми, определить их специфику и перспективы. В случае интеграции с IDS основные надежды российских исследователей связаны с изучением естественного движения населения и форм эволюции семьи, а также специфики этноконфессиональных сообществ.

Помимо российского, на общем собрании EHPS-Net было заявлено о начале еще двух проектов баз данных: Норвежского исторического регистра населения (Университет Тромсё) и Исторической базы населения Трансильвании (Университет Клуж, Румыния).

В Альгеро отмечалось в целом удовлетворительное состояние реализации проекта EHPS-Net. Узким местом сегодня является разработка программного обеспечения, в связи с чем решено подключить сторонних разработчиков для ускорения работы. Тем не менее уже в ближайшее время предполагается выложить тестовые образцы интерфейсных программ и программного обеспечения IDS на сайте проекта.

Важной стороной деятельности EHPS-Net является система профессиональной подготовки специалистов. Она включает несколько уровней. Обучение начинающих осуществляется на базе университета в Клуже (Румыния). Обычный состав группы — 20 человек, хотя высказываются мнения о необходимости увеличения числа слушателей до 40. Студенты овладевают базовыми навыками работы с историко-демографическими данными, для чего требуется предварительное знание ими программных статистических пакетов *Stata* и *SPSS*. Часть расходов на размещение и питание слушателей компенсируется организаторами. Основная проблема для привлечения большего числа слушателей из стран Восточной Европы, в том числе из России, — языковой барьер (занятия идут на английском языке). Тем не менее, в работе школы в 2014 г. принимали участие несколько молодых исследователей из Барнаула.

«Продвинутый» уровень предлагают школы в Лёвене (Бельгия), Умеа и Лунде (Швеция), а также в Амстердаме и Женеве. Как правило, они проводятся раз в два года и учат работе с IDS, а также с местными базами демографических данных.

Несмотря на незавершенность проекта, он уже принес результаты, в первую очередь в изучении истории семей, а также небольших сообществ. Так, по ряду стран расширен объем информации в отношении отдельных лиц путем совмещения данных церковного и налогового учетов; появилась возможность реконструкции индивидуальных семейных стратегий, что позволяет изучать межпоколенные сценарии передачи социального статуса и культурных традиций; инициированы исследования на больших временных отрезках, имеющие целью сравнение социально-экономического и демографического поведения мигрантов и немигрантов, а также горожан, жителей пригородов и села. Кроме того, в современных европейских демографических исследованиях в большей степени, чем прежде, присутствует сравнительный анализ, опирающийся не только на европейские данные.

Проект EHPS-Net позволяет говорить о появлении сравнительной исторической демографии в ее синхронном и диахронном восприятии. Многие проблемы, с которыми столкнулся современный мир (приватизация семейной жизни, ослабление групповой идентичности, переоценка роли женщин и их положения в семье и обществе), имели место уже во второй половине XIX столетия. Сравнительные историко-демографические данные позволят понять перспективы развития современного мира, истоки имеющихся различий в стратегиях брачности, семейного развития, миграционной мобильности, характере смертности.

Предполагается, что на новый уровень будут выведены миграционные исследования. Сравнительный анализ миграций поможет оценить перспективы приживаемости мигрантов в крупных городах, особенности жизненных стратегий мигрантов второго и третьего поколений.

Представители EHPS-Net приняли активное участие в организации Европейского общества исторической демографии (European Society of Historical Demography, ESHD[7]). Оно было создано усилиями трех национальных историко-демографических сообществ — Испании, Италии и Франции, а учредительная конференция ESHD также была созвана в Альгеро в сентябре 2014 г. Деятели EHPS-Net составили примерно половину Совета вновь созданного общества. На конференции было представлено около 130 докладов участников из большинства стран Европы, а также из Бразилии, Канады, США, Японии. К сожалению, российская делегация была крайне немногочисленной, в ней были упомянутые исследователи из Уральского федерального университета, с докладами выступили Е. М. Главацкая («Католическая община на Среднем Урале: жизненные циклы (по материалам метрических книг конца XIX — начала XX вв.») и Л. Н. Мазур («Модели демографической эволюции сельских поселений под влиянием урбанизации (по материалам Среднего Урала в XX в.)»).

Помимо многочисленных секций, посвященных конкретным проблемам исторической демографии (детства, рождаемости, брака, смертно-

сти), ожидаемо нашлось место вопросам обработки и интерпретации историко-демографических данных. На одной из секций обсуждались особенности и перспективы использования первичных данных. Была представлена франко-итальянская компьютерная программа для определения степени родства на основе доступных персональных сведений. Состоялась также презентация электронного журнала EHPS-Net «Historical Life Course Studies».

Было решено, что конференция будет проводиться раз в два года, и следующее подобное мероприятие должно пройти в 2016 г. в Бельгии.

Стремление европейских историков народонаселения к интеграции пока не встречает массового отклика у их российских коллег. Одним из оправданий может служить недостаток первичной источниковой информации в российских архивах. Но, как показывает опыт ученых Москвы, Санкт-Петербурга, Барнаула, Тамбова и Екатеринбурга, эти проблемы преодолимы. Сравнительные исследования с использованием российского материала необходимы для того, чтобы вывести отечественную историко-демографическую науку на новый методологический и содержательный уровень.

ПРИМЕЧАНИЯ

1 Minnesota Population Center [Electronic resource]. URL: http://www.ipums.org.

2 European Historical Population Samples Network [Electronic Resource]. URL: http://www.ehps-net.eu/

3 В январе 2015 г. появился второй выпуск журнала (см.: European Historical Population Samples Network. Historical Life Course Studies [Electronic resource]. URL: http://www.ehps-net.eu/journal).

4 European Historical Population Samples Network. Database Map [Electronic resource]. URL: http://www.ehps-net.eu/content/database-map

5 European Historical Population Samples Network. IDS [Electronic resource]. URL: http://www.ehps-net.eu/content/ids-repository

6 См. подробнее: Alter, G. & Mandemakers, K. (26–05–2014). The Intermediate Data Structure (IDS) for Longitudinal Historical Microdata, version 4. Historical Life Course Studies, Volume 1, 1–26 [Electronic resource]. URL: http://hdl.handle.net/10622/23526343–2014–0001? locatt=view: master.

7 European Society of Historical Demography [Electronic resource]. URL: http://www.eshd.eu.