

Е.А. Брюханова**Переписи населения в информационном пространстве:
опыт обработки профессиональных данных переписи 1897 г. ***

Широкое применение информационных технологий в исторических исследованиях отражает прочно утвердившиеся тенденции информатизации общества в целом. Это отражается не только в расширении методов обработки и анализа данных различного вида, но и в способах сбора и представления информации исторических источников. В качестве примера информатизации исторических исследований можно привести проекты межнационального, российского и регионального уровней, направленные на сбор, обработку и анализ данных переписей населения. Особый интерес для нашего исследования представляют данные Всероссийской переписи населения 1897 г. Это обусловлено, во-первых, комплексностью собираемых данных и масштабностью статистического обследования, впервые проведенного на конкретную дату и по всей территории страны, во-вторых, тем, что предоставляемые переписью сведения являются востребованными исследователями разных наук, при этом актуальным становится вопрос доступности не столько первичных (т.к. остаются не решенной проблемы сохранения таких источников), сколько агрегированных данных, что подразумевает создание открытых информационных систем, расширяющих исследовательские возможности.

Переписи населения, вне зависимости от уровня, целей, задач и времени проведения, являются одними из самых востребованных источников, актуальных для исследователей разных периодов и направлений. Комплексность данных, предоставляемых переписями и другими статистическим обследованиями, объясняет значительное внимание исследователей, с одной стороны, к сохранению, а с другой, – к расширению доступа к таким материалам, что открывает широкий спектр новых исследовательских возможностей и перспектив. Так, в контексте развития исторической демографии в ряде европейских стран реализовывались проекты по созданию и опубликованию персональных данных, полученных в ходе национальных переписей населения. Например, на сайте национального архива Норвегии выложена информация первичных переписных листов 20 национальных переписей с 1801 г. по настоящее время¹. Существуют и крупные межнациональные проекты, ориентированные на сбор, обработку и создание баз данных на основе персональных сведений исторических источников. К таким проектам относятся:

Брюханова Елена Александровна, к.и.н., доцент, Алтайский государственный университет. Эл. почта: brukhanova@hist.asu.ru

** Исследование выполнено при поддержке РФНФ, проект № 11-01-00455а*

¹ Расцвет исторической информатики. Леонид Бородкин о клиодинамике и моделировании прошлого// НГ Ex libris. №18. 26.05.2011 [Электронный ресурс]. Режим доступа: <http://mikhail-boyko.narod.ru/interview/borodkin.html>; Digitalarkivet [Электронный ресурс]. Режим доступа: <http://www.arkivverket.no/eng/content/view/full/629>

- North Atlantic Population Project – проект по изучению населения североатлантического региона по данным переписей населения, проведенных в Канаде, Великобритании, Германии, Исландии, Норвегии, Швеции и США с 1810 по 1910 г.¹

- IPUMS (Integrated Public Use Microdata Series) – проект, направленный на аккумуляцию, интеграцию и опубликование данных переписей населений всего мира для социальных и экономических исследований. В проекте приняли участие 68 стран, которые предоставили данные по 211 переписям, включающим 480 миллионов персональных записей².

- MOSAIC project – проект, предусматривающий объединение и дополнение уже существующих массивов, включение новых персональных данных переписей населения с целью проведения международных сравнительных исследований вне времени и пространства³.

В России на настоящий момент не реализуется ни одного проекта по сбору и обработке персональных данных хотя бы одной переписи населения по всей её территории. Объясняется это обширностью территории и плохой сохранностью первичных материалов переписей. Но при этом в России разными исследовательскими коллективами создаются базы персональных данных по переписям и другим статистическим материалам отдельных регионов. Например, по переписи 1897 г. были созданы базы данных по населению городов Тобольска и Ярославля, существует база данных по 5% выборке сельскохозяйственной переписи 1917 г. в Алтайской губернии, продолжают пополняться базы данных по метрическим книгам городов Барнаула, Ярославля, Севастополя. Архивные учреждения также включаются в процесс информатизации и предлагают исследователям базы данных, в том числе и статистических обследований. Так, Государственный архив Алтайского края открыл доступ для посетителей читального зала к базе данных «Анкеты Всероссийских переписей 1917 года»⁴. Это только некоторые примеры статистических выборок, как правило, ориентированных на выполнение какого-либо проекта, а поэтому чаще не источников-ориентированных, а тематических, т.е. они могут не содержать часть информации источника о социальном положении или занятии.

Таким образом, для комплексных межрегиональных исследований особое значение приобретают не персональные, а агрегированные данные переписей и других статистических обследований. При этом остро ощущается потребность в доступности таких данных и возможности их адекватной обработки. Следует отметить, что в информационном пространстве сети интернет можно найти значи-

¹ North Atlantic Population Project [Электронный ресурс]. Режим доступа: <http://www.nappdata.org/napp/>

² Integrated Public Use Microdata Series, International. Census Microdata for Social and Economic Research [Электронный ресурс]. Режим доступа: <https://international.ipums.org/international/>; IPUMS: Международный проект по обеспечению открытого доступа к коллекциям микроданных [Электронный ресурс]. Режим доступа: <http://www.aik-sng.ru/text/bullet/29/29.html>

³ Recovering Surviving Census Records to Reconstruct Population, Economic, and Cultural History [Электронный ресурс]. Режим доступа: <http://www.censusmosaic.org/cgi-bin/index.plx>

⁴ Перечень автоматизированных баз данных, доступных исследователям в читальном зале КГКУ «Государственный архив Алтайского края» [Электронный ресурс]. Режим доступа: http://www.archiv.ab.ru/fond_info/perechen_el_res/1_1/

тельный объем оцифрованных статистических источников, в том числе по переписи 1897 г., на сайтах научных библиотек или специализированных порталах ¹. Но формат, в котором предоставляется информация, не позволяет оперативно использовать статистические ряды данных. В этом контексте появляются проекты, направленных на перевод статичной информации в статистические таблицы или реляционные базы данных, позволяющие полноценно использовать материалы источника. Так, в настоящее время реализуется проект по созданию «Электронного архива российской исторической статистики, XVIII–XXI вв.», целью которого является сбор и организация в виде электронной базы данных, опубликованной в сети интернет, исторической информации из различных опубликованных и неопубликованных источников. Доступ к электронному архиву планируется открыть в сентябре 2013 г.²

Всероссийская перепись населения 1897 г. – одна из наиболее масштабных статистических кампаний Российской империи конца XIX – начала XX в. Она охватила всю территорию страны и представила единовременный срез данных о населении по целому ряду признаков, в том числе об основных и побочных занятиях. Учитывая объем и качество собранной переписчиками информации, материалы переписи 1897 г. являются широко востребованными в исторических исследованиях. Данное обстоятельство обусловило актуальность проектов, направленных на создание информационных систем, основанных на агрегированных данных переписи 1897 г. Один из таких проектов был реализован в Лаборатории социальной истории Тамбовского государственного университета при поддержке Российского гуманитарного научного фонда. В рамках проекта планировалось разработать, обобщить, апробировать методику создания и публикации баз данных по социально-демографической истории в Интернет на примере создания базы данных по материалам Первой всеобщей переписи населения Российской империи 1897 г.³ Создание такой электронной информационной системы должно было решить проблему доступности большого количества агрегированных данных переписи 1897 г., предоставив пользователям удобный интерфейс базы данных и открытый доступ к ней в сети Интернет. К сожалению, ограничение доступа к ресурсу через систему неавтоматической регистрации не позволило в полной мере реализовать заявленные задачи.

В настоящее время на историческом факультете Алтайского государственного университета при поддержке Российского гуманитарного научного фонда реализуется проект, основанный на агрегированных данных переписи 1897 г., отражающих профессии и занятия населения Российской империи в конце XIX в. В ходе

¹ Докусфера: Электронный фонд Российской национальной библиотеки [Электронный ресурс]. Режим доступа: <http://leb.nlr.ru/>; Президентская библиотека им.Б.Н. Ельцина [Электронный ресурс]. Режим доступа: <http://www.prilib.ru/Lib/pages/catalog.aspx?catid=2255>; Утерянная империя. Статистика [Электронный ресурс]. Режим доступа: http://lost-empire.ru/index.php?option=com_content&task=view&id=23&Itemid=9; Переписи населения на Демоскоп Weekly [Электронный ресурс]. Режим доступа: <http://demoscope.ru/weekly/ssp/census.php?cy=0>

² Электронный архив российской исторической статистики, XVIII–XXI вв. [Электронный ресурс]. Режим доступа: http://www.dynastyfdn.com/programs/education/electronic_repository

³ Первая перепись населения Российской империи 1897 г. [Электронный ресурс]. Режим доступа: <http://www.census1897.com/?view=00#>

переписи 1897 г. впервые сведения об основном и побочном занятии населения собирались и обрабатывались по единой программе в масштабе всей страны. Определенные особенности источника, такие как большое количество профессиональных групп, требующих дополнительной обработки, критические замечания статистиков, экономистов, историков и др. специалистов, например, замечание В.И. Ленина о том, что «статистика эта, по всей видимости, грешит весьма значительным преуменьшением числа рабочих»¹, ограничили использование данных о профессиональной занятости населения в исторических исследованиях. Проект направлен на актуализацию исследований в области профессиональной занятости населения конца XIX в. на основе данных переписи 1897 г., а также создание геоинформационной системы для проведения пространственно-географического анализа профессиональной структуры населения России. Особенностью проекта является применение международных систем классификации занятий (HISCO, PST) для построения моделей социально-профессиональной структуры российского общества. Реализацию проекта можно представить в несколько этапов:

1. определение состава и структуры источников, предварительный источниковедческий и документоведческий анализ опубликованных агрегированных данных;
2. перевод данных в формат статистических таблиц, разработка структуры базы данных, создание дополнительных справочников;
3. кодирование профессиональных групп переписи 1897 г. в системы международных классификаций занятий – HISCO, PST;
4. создание геоинформационной системы по занятиям населения переписи 1897 г.

Работа с агрегированными профессиональными данными Первой всероссийской переписи населения 1897 г. началась с анализа её результатов, публикуемых Центральным статистическим комитетом. Данные о профессиях и занятиях населения Российской империи публиковались в нескольких изданиях: «Общем сводке по империи результатов разработки данных первой всеобщей переписи населения, произведенной 28 января 1897 г.» (в двух томах), 89 выпусках по губерниям и областям Российской империи, четырем городам (Петербург, Москва, Одесса, Варшава) и острову Сахалин, в специальном издании «Распределение населения по видам главных занятий и возрастным группам по отдельным территориальным районам» (в четырех томах), в издании «Численность и состав рабочих в России на основании данных Первой всеобщей переписи 1897 г.» (в двух томах). В совокупности агрегированные профессиональные данные переписи 1897 г. предоставляют следующие сведения: пол, возраст, национальность занятых, территориальное распределение занятий (по переписным и территориальным районам, губерниям, округам (уездам), городам), для рабочих и прислуги – грамотность, семейное положение и состав семьи. Профессиональная структура представлена в виде 65 групп и 360 видов профессиональной деятельности. Сводная таблица по изданию ЦСУ и содержащимся в них характеристикам профессиональных групп представлена в Приложении 1.

В настоящее время созданная информационная система по агрегированным профессиональным данным переписи 1897 г. основана на материалах 50 губерний Европейской России и девяти губерний Сибири, и включает следующие характери-

¹ Ленин В.И. Полн. собр. соч. Т. 3. С. 502.

стики профессиональных групп: административно-территориальное деление, пол, возрастные группы, народности и количество членов семьи. Набор характеристик (например, количество или процент грамотных) может расширяться по мере дополнения системы новыми данными.

При проектировании структуры базы данных важным было, во-первых, обеспечение привязки к источнику, т.е. конкретному изданию ЦСУ, что позволит сопоставлять данные разных выпусков результатов переписи и провести комплексный источниковедческий анализ; во-вторых, сохранение структуры таблицы, аналогичной источнику; в-третьих, возможность свободной компиляции данных, схожих по структуре (т.е. по количеству профессиональных групп), что обеспечивает удобство работы с данными и возможность проведения многомерного профессионального анализа.

Одновременно с формированием структуры базы данных создаются дополнительные справочники: справочник административно-территориальных единиц (губернии, уезды, города), справочник народностей Российской империи конца XIX в., справочник профессиональных групп с атрибутами в виде кодов международных систем классификации HISCO и PST.

Суть применения международных классификационных систем заключается в кодировании профессиональных групп переписи, принцип организации которых исследователями (Б.П. Кадомцевым, Г.Г. Швиттау, Б.Н. Мироновым и др.) был признан неудачным и не позволяющим сопоставлять их с данными других источников. Код, выступая универсальным средством отражения информации, позволяет включить агрегированные профессиональные данные переписи 1897 г. в сравнительно-исторические исследования, в том числе и на международном уровне.

Одним из перспективных направлений исторических исследований представляется применение геоинформационных технологий в изучении профессионального состава, структуры и профессиональной мобильности. В настоящее время создается геоинформационная система «Профессии и занятия населения Европейской России и Сибири конца XIX – начала XX в.», основанная на агрегированных профессиональных данных переписи 1897 г. Она включает набор тематических карт территориального распределения отдельных профессиональных групп по различным социально-демографическим характеристикам: пол, возраст, народность и т.д. Особое внимание уделяется отражению в геоинформационной системе международных классификаций занятий населения как представление профессиональной структуры того или иного региона. В качестве примера успешного соединения геоинформационных технологий и метода классификации профессий можно привести проект Кембриджского университета по профессиональной структуре Британии 1379–1911 гг., представленной в классификации PST¹. В России аналогичных проектов с использованием ГИС-технологий не проводилось. Создание такой системы позволит провести сравнительный пространственно-географический анализ профессионального состава населения различных регионов России, а именно выделить региональную и отраслевую структуру и специфику занятости населения, выявить отраслевые связи и межрегиональные профессиональные корпоры.

¹ The occupational structure of Britain 1379-1911 [Электронный ресурс]. Режим доступа: <http://www.geog.cam.ac.uk/research/projects/occupations/>

Таким образом, данные переписей населения (как агрегированные, так и персональные), являясь объектом повышенного внимания со стороны специалистов разных наук, всё больше включаются в процессы информатизации, оцифровки и «вживления» в мировое виртуальное пространство. И, если за рубежом можно увидеть крупные проекты, интегрирующие большое количество микроданных, т.е. отдельных переписей и статистических обследований разных периодов, то для России характерно точечная тематическая разработка уникальных источниковых ресурсов, например, переписных листов переписи 1897 г. Обширность территории и плохая сохранность персональных данных обусловили актуальность разработки информационных систем на основе агрегированных статистических материалов. Создание таких систем значительно расширяет диапазон исторических исследований, представляя возможность использования в том числе и геоинформационных технологий.

Приложение 1.

**Источники данных о профессиях и занятиях населения
по переписи 1897 г.**

Название	Территория	Кол-во проф. групп	Сведения
Общий свод по империи результатов разработки данных первой всеобщей переписи населения, произведенной 28 января 1897 г.	Территориальные районы: Европейская Россия Привислинская губерния Кавказ Сибирь Средняя Азия Санкт-Петербург Москва	65 групп 360 видов	- самодеятельное население (по полу); - члены семей (по полу)
Первая всеобщая перепись населения Российской империи, 1897. По губерниям	Губерния Уезды Города	65 групп	Самодеятельное население и члены семей: - по полу - по возрастным группам - по административным единицам - по народностям - по побочным промыслам
Распределение населения по видам главных занятий и возрастным группам по отдельным территориальным районам	Переписной округ Губерния Уезд Крупнейшие города	65 групп 360 видов	Самодеятельное население: - по полу - возрастным группам (количество занятых по переписному округу) Членов семей по полу в данных по губернии
Численность и состав рабочих в России на основании данных Первой всеобщей переписи 1897 г.	Территориальным районам Губерниям	28 групп 149+10 видов	Рабочие: - по полу - по губерниям - отдельно лица, моложе 15 лет - по возрастным группам - по семейному положению и составу семьи - по грамотности