

УДК 930.25

doi 10.17072/2219-3111-2024-4-69-83

Ссылка для цитирования: Брюханова Е. А., Иванова Н. П. Базы данных и источниковедческий анализ в цифровую эпоху: возможности и ограничения (на примере данных Переписи 1897 года по Сибири) // Вестник Пермского университета. История. 2024. № 4(67). С. 69–83.

БАЗЫ ДАННЫХ И ИСТОЧНИКОВЕДЧЕСКИЙ АНАЛИЗ В ЦИФРОВУЮ ЭПОХУ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ (НА ПРИМЕРЕ ДАННЫХ ПЕРЕПИСИ 1897 ГОДА ПО СИБИРИ)¹

Е. А. Брюханова

Алтайский государственный университет, 656049, Россия, Барнаул, пр-т Ленина, 61

brelenochka@mail.ru

ResearcherID: N-9527-2016

Scopus Author ID: 57194780366

SPIN-код: 1342-7026

Н. П. Иванова

Алтайский государственный университет, 656049, Россия, Барнаул, пр-т Ленина, 61

ivanovanp@gmail.com

ResearcherID: JPK-0593-2023

SPIN-код: 5829-2234

Современный этап развития исторической науки и гуманитарных наук в целом характеризуется активным созданием цифровых проектов и ресурсов, в рамках которых формируются новые виды источников: оцифрованные документы, базы данных, информационные системы и т.п. Подобные тенденции расширяют возможности исследователей, а также ставят перед ними новые задачи и открывают новые направления. Одним из таких актуальных направлений является источниковедение информационных ресурсов и цифровых данных. Особый интерес новые инструменты представляют для исторических источников, которые носят дискуссионный характер, например, материалы Первой всеобщей переписи населения Российской империи 1897 г. Цель исследования заключается в оценке возможностей использования технологии баз данных для источниковедческого анализа материалов Переписи 1897 г. на примере данных по численности и занятости Сибири. В статье решаются две взаимосвязанные задачи: рассмотреть преимущества и ограничения баз данных для источниковедческого анализа материалов Переписи 1897 г. и оценить достоверность ее данных по отдельным показателям. Источниками статьи послужили как собственно опубликованные и архивные материалы Переписи 1897 г. (агрегированные данные и персональные сведения переписных листов по городу Тобольску), так и базы данных, созданные по ним. В основу исследования был положен системный подход, а для источниковедческого анализа использовались ретроспективный, сравнительный методы и внутренняя критика источников. Проведенное исследование показало широкие возможности баз данных для источниковедческого анализа и позволило сделать вывод о высоком уровне сбора и обработки материалов Переписи 1897 г. по населению Сибири: разнотченность по численности города Тобольска по агрегированным и персональным данным составило 1 %. Кроме того, развитие цифровых ресурсов и доступность для пользователей материалов Переписи 1897 г. и баз данных не только расширяет возможности исследователей, но и обеспечивает верификацию данных и выводов.

Ключевые слова: источниковедческий анализ, статистические материалы, переписные листы, база данных, Перепись 1897 г., женская занятость, Тобольск, Сибирь.

Введение. Исторические источники нового типа и вопросы источниковедческого анализа

Наука всегда реагировала на изменения, происходящие в обществе, а зачастую сама создавала актуальные тренды. В истории одним из таких актуальных трендов были и остаются информационные технологии, которые давно уже вошли в практику и инструментарий

исторических исследований [Гарскова, 2018; Владимиров, 2006; Корниенко, 2021, с. 24–25]. Современный период проходит под знаком «цифрового поворота» [Бородкин, 2019] не только в истории, но и в гуманитарных науках в целом [Можжаева, 2015]. В активно развивающемся направлении Digital Humanities (цифровые гуманитарные науки) особое значение придается междисциплинарности в проектах. Исследователи видят в этом расширение возможностей для изучения прошлого в цифровом пространстве с помощью соответствующих инструментов и медиа [Володин, 2014].

Одной из ключевых особенностей нового этапа является изменение подходов к определению понятия «данные», их поиску, аккумуляции, хранению и анализу, а также репрезентации в цифровой среде. В практике зарубежных исследований уже используются как крупные базы по исторической статистике [Historical Statistics...], так и международные банки лонгитюдных данных по сведениям регистров, переписей и других форм учета населения, представленные на проектах *IPUMS (Integrated Public Use Microdata Series)* [Ruggles, p. 287] и *Minina Microdata: Economic Opportunity and Spatial Mobility in Britain, Canada and The United States, 1850–1911* [Inwood, 2014]. В России столь масштабные инфраструктурные проекты в области исторических данных – дело будущего. Тем не менее можно отметить некоторые успешно реализованные проекты. Например, «Динамика экономического и социального развития России в XIX – начале XX вв.» [Динамика экономического и социального развития...] и «Электронный архив Российской исторической статистики (Ристат)» [Ристат...] предоставляют пользователю структурированные и готовые для дальнейшего использования данные. Более того, реальностью исторических исследований стало использование коллекций оцифрованных исторических документов и материалов библиотек и архивов для создания собственных ресурсов с массивами цифровых данных. Именно так создавались базы данных и информационные системы по материалам Первой всеобщей переписи населения Российской империи 1897 г., в которых представлены как статистические данные, так и сведения из переписных листов.

Можно отметить, что в эпоху цифрового поворота активно создаются исторические источники нового типа и формата, предоставляющие исследователям не только широкие возможности, но и формирующие новые вызовы. Один из актуальных вызовов находится в сфере источниковедения информационных ресурсов и цифровых данных.

Исследователи все чаще отмечают проблемы «источниковедения цифровых документов» [Бородкин, 2019, с. 19–20], те аспекты, которые следует учитывать пользователям информационных ресурсов, например, критика оцифрованных источников [Володин, 2014, с. 7; Юмашева, 2018], оценка релевантности, репрезентативности и достоверности цифровых данных [Корнев, 2018, с. 64]. Основные этапы развития источниковедения традиционных и электронных исторических источников, проект классификации последних и методы их источниковедческого анализа предложены Ю. Ю. Юмашевой [Юмашева, 2023, с. 113–114, 123].

Большая часть отмеченных выше аспектов применения цифровых проектов зависит от качества источниковедческого анализа на этапе разработки и корректного формирования метаданных. Предпроектный анализ источников направлен не только на разработку структуры проекта, создание необходимых внутренних элементов (справочников, таблиц, форм ввода и т.д.), выстраивание логики данных, но и на сохранение связей «источник – данные», что позволяет пользователям проводить верификацию, оценивать их релевантность и репрезентативность для собственного исследования.

Источниковедческий анализ может стать самостоятельной задачей при использовании уже готового информационного ресурса или цифрового проекта, позволяя исследователю получить новую информацию об источниках, например, выявить ошибки, пропуски или особенности структуры данных, влияющие на их использование, провести реконструкцию сбора или обработки исторических сведений, оценить достоверность источника и др. Такие задачи могут решаться как на этапе создания проекта (если они не являются основной целью), так и стать отдельным направлением исследования в период использования информационного

ресурса. Особый интерес представляют постпроектные исследования для исторических материалов, источниковедческая оценка которых относится к дискуссионным вопросам.

К таким критически оцениваемым источникам можно отнести материалы Первой всеобщей переписи населения Российской империи 1897 г. Несмотря на то что ее данные являются одними из наиболее востребованных в исторических исследованиях, такие источниковедческие вопросы, как, например, масштабность публикаций ее результатов [Акашева, 2023], выявление и структура первичных материалов [Брюханова, Иванова, 2019], этого самого грандиозного статистического мероприятия конца XIX в. до сих пор остаются актуальными.

Итак, целью исследования является оценка возможностей технологии баз данных для источниковедческого анализа материалов Первой всеобщей переписи населения Российской империи 1897 г. на примере данных по Сибири.

В статье авторы предлагают решить две взаимосвязанные задачи:

– рассмотреть преимущества и ограничения баз данных для источниковедческого анализа материалов Переписи 1897 г.;

– оценить достоверность ее данных по отдельным показателям на основе компаративного анализа сведений информационных ресурсов.

Источники

Под источниками авторами понимаются как собственно материалы Первой всеобщей переписи населения 1897 г., так и базы данных, созданные по ним. Материалы Переписи 1897 г. в настоящем исследовании предусматривают два уровня данных: агрегированные (или статистические) по губерниям Сибири и первичные, а именно персональные сведения из переписных листов по городам Тобольской губернии. На основе материалов Переписи 1897 г. были созданы две реляционные источник-ориентированные базы данных и их онлайн-версии, представленные в информационных системах.

База данных «Профессии и занятия населения Российской империи конца XIX – начала XX веков (по материалам Всероссийской переписи населения 1897 года)» (далее – база данных «Профессии и занятия») содержит сведения четырех таблиц о профессиях, занятиях и структуре занятого населения регионов, губерний, уездов и городов Российской империи (Профессии и занятия населения..., 2014), опубликованных в Общем свode результатов переписи (Общий свод..., 1905), погубернских итогах (Первая Всеобщая перепись..., 1905) и специальном издании по занятиям населения (Распределение населения..., 1905). Основной концепцией базы данных стала декомпиляция достаточно жестких статистических таблиц и их интеграция в систему управления базами данных MySQL, которая позволяет хранить (кроме чисел из источника) вычисленные значения и формировать многомерные отчеты, расширяющие изначальные возможности источника. Такой подход позволил провести внутреннюю критику данных отдельных выпусков результатов переписи, а также осуществить сопоставление сведений по трем источникам.

База данных «Население городов Сибири на рубеже XIX–XX вв.: по материалам Первой всеобщей переписи населения Российской империи 1897 г.» (Население городов Сибири..., 2021) (далее – база данных «Население городов») и ее онлайн-версия (Население городов Сибири на рубеже XIX–XX вв. по материалам Первой всеобщей переписи населения 1897 г.) включают сведения из ведомостей, обложек и переписных листов по городам Сибири. База данных заполнялась по принципу сплошной выборки по сохранившимся переписным листам. В настоящем исследовании будут использоваться персональные данные из переписных листов по г. Тобольску (ГАТ. Ф. 417. Оп. 2. Д. 1–38), что обусловлено их полной сохранностью. В базу данных вносились все сведения из переписного листа: исправленные и дописанные сведения, вычеркнутые записи. Такой подход на основе ретроспективного метода и критики данных позволяет проследить этапы сбора и корректировки материалов переписи. Компаративный анализ двух баз данных обеспечит реконструкцию методики группировки отдельных данных и, возможно, оценку достоверности опубликованных результатов переписи.

Источниковедческий анализ агрегированных данных Переписи 1897 г.

Приступая к источниковедческому анализу материалов Переписи 1897 г. с использованием технологии баз данных, мы принимаем во внимание, что созданные на основе опубликованных и архивных материалов переписи базы данных являются самостоятельными метаисточниками, которые имеют как преимущества, так и некоторые ограничения. Например, база данных «Профессии и занятия» позволяет рассматривать весь комплекс опубликованных результатов переписи как систему взаимосвязанных данных, элементы которой не имеют жестких границ и могут компоноваться между собой, исходя из исследовательских задач. Представленная система ограничена несколькими векторами с разным уровнем данных, заложенными в самом источнике, такими как:

- административно-территориальное деление разного уровня (Российская империя – регион – губерния – уезд – город);
- категории населения (городское/сельское/общее; самозанятые/члены семей; мужчины/женщины);
- группировка профессиональной занятости (65 групп и 360 видов);
- дополнительные характеристики занятых и членов семей (возраст, народности, грамотность и др.), которые, в свою очередь, имеют свои подгруппы.

Отсутствие же в базе данных жестких границ позволяет получить:

- срезы данных, отсутствующие в опубликованных результатах переписи, например, численность населения произвольно выбранной территории, составленной из уездов, сведения о которых могли находиться в разных выпусках и таблицах, что может быть полезно при изучении современных регионов;
- агрегацию данных из разных выпусков до более высоких уровней административного деления (Российская империя – регион), не предусмотренных по отдельным изданиям, что дает возможность сопоставления сведений из разных выпусков и таблиц;
- вычисленные значения (по административным единицам, группам занятий и категориям населения), что позволяет не только провести внутреннюю критику и выявить ошибки первоисточника, но и осуществить реконструкцию работы счетчиков при получении агрегированных данных и определить верные значения этих данных.

В нашем исследовании методика проверки агрегированных данных Переписи 1897 г. по Сибири проходила в два этапа. На первом этапе с помощью запроса были сопоставлены: 1) итоговые значения по численности населения Сибири в целом и восьми ее губерниям по трем источникам (Общий свод..., 1905; Первая Всеобщая перепись..., 1905; Распределение населения..., 1905); 2) показатели занятости населения и членов семей по 65 группам по сибирским губерниям по двум источникам (Общий свод..., 1905; Первая Всеобщая перепись..., 1905).

Результаты сопоставления численности населения Сибири по разным источникам и таблицам Переписи 1897 г. представлены в табл. 1, из которой видно, что разночтения между источниками, представленными разными выпусками, статистически незначительны и составляют десятитысячные доли процента.

Второй этап заключался в сравнении данных двух источников (Общий свод..., 1905; Первая Всеобщая перепись..., 1905) по категориям населения (самозанятые / члены семей) и группам занятий на уровне Сибири в целом и губерний. Результаты сопоставления показали, что статистически значимые разночтения между источниками несущественны и не влияют ни на показатели численности населения (регион – губерния), ни на соотношение самозанятых и членов семей, ни на численность занятых в профессиональных группах (по 65 группам). Тем не менее выявленные разночтения оказались типичными для погубернских изданий и их можно обобщить следующим образом.

В погубернских выпусках в некоторых случаях происходил перенос членов семей в категорию самозанятого населения: например, в группе 09 «Учебная и воспитательная деятельность» по Томской губернии 14 человек мужчин и 3 женщины из категории «Самозанятые» в Общем своде переместились в категорию «Члены семей» в выпуске по губернии (табл. 2).

Таблица 1

Сопоставление численности населения Сибири по разным выпускам результатов Переписи 1897 г.²

Численность населения	Общий свод, данные по Сибири (табл. XX-а)	Общий свод, агрегация по сибирским губерниям (табл. XX-а)	Агрегация из погубернских выпусков (табл. XXI)	Агрегация из таблиц по губерниям Сибири из специального издания по занятиям (табл. XX)
Сибирь	5 758 822	5 758 829	5 758 822	5 758 828
Губернии				
Амурская		120 306	120 306	120 306
Енисейская		570 161	570 161	570 161
Забайкальская		672 037	672 037	672 037
Иркутская		514 267	514 267	514 267
Остров Сахалин		28 113	28 113	28112
Приморская		223 336	223 336	223 336
Тобольская		1 433 043	1 433 043	1 433 043
Томская		1 927 686	1 927 679	1927 686
Якутская		269 880	269 880	269 880

Примечание: выделены ячейки с отличающимися значениями по губерниям.

В некоторых погубернских выпусках самозанятые и/или члены семей были перенесены из одной группы в другую. Например, по Енисейской губернии 55 женщин – членов семей, учтенных в Общем своде в группе 06 «Богослужение других христианских исповеданий», в выпуске по губернии оказались в группе 05 «Богослужение православного исповедания» (см. табл. 2).

Предполагаем, что могли присутствовать и ошибки счета данных в группах 04 «Вооруженные силы» и 15 «Живущие на счет казны и общественных учреждений и на средства частных лиц», в которых разночтения встречаются в нескольких губерниях.

Таблица 2

Разночтения данных по группам занятий по двум источникам³

Группы занятий	Общий свод (табл. XX)				Разночтения с данными, полученными путем агрегации сведений из погубернских выпусков				
	Самозанятых		Членов семей		Самозанятых		Членов семей		Губернии, в которых выявлены разночтения по группам (чел.)
	м	ж	м	ж	м	ж	м	ж	
Сибирь в целом	1 367 308	214 430	1 597 118	2 579 973	-13	-70	+20	+70	
01. Администрация, суд и полиция	9440	40	4951	11 402	-1	-6	+1	+6	Сахалин (1), Енисейская (5)
04. Вооруженные силы	59 408	1	1116	2763	-2 6	0	0	0	Томская (2), Тобольская (8), Иркутская

Группы занятий	Общий свод (табл. XX)				Разночтения с данными, полученными путем агрегации сведений из погубернских выпусков				
	Самозанятых		Членов семей		Самозанятых		Членов семей		Губернии, в которых выявлены разночтения по группам (чел.)
	м	ж	м	ж	м	ж	м	ж	
									(5), Забайкальская (3), Иркутская (5), Приморская (8)
05. Богослужение православного исповедания	4635	1735	3503	8052	0	-4	0	+59	Томская (4) Енисейская (55)
06. Богослужение других христианских исповеданий	33	12	17	23	0	0	0	-55	Енисейская (55)
09. Учебная и воспитательная деятельность	3779	1907	1478	3048	-1 4	-3	+1 4	+3	Томская (вся строка)
15. Живущие на счет казны и общественных учреждений и на средства частных лиц	17 615	13 063	717	1248	33	0	0	0	Томская (9), Тобольская (8), Забайкальская (3), Иркутская (5), Приморская (8)
38. Работы по устройству, ремонту и содержанию жилищ и строительные вообще	55 773	94	20 768	43 928	0	-5 7	0	+57	Томская (43), Енисейская (6), Забайкальская (4), Иркутская (4)
43. Извозный промысел	13 240	162	7386	13 917	-5	0	+5	0	Томская (1), Тобольская (4)
44. Остальные сухопутные сообщения и средства передвижения	699	46	235	492	-1	-1	0	-7	Сахалин
45. Почта, телеграф и телефон	2108	78	1132	2627	+1	+1	0	+7	Сахалин
58. Торговля остальными предметами	672	25	381	810	0	-1	0	0	Сахалин
60. Трактиры, гостиницы, меблированные комнаты, клубы	2359	931	1415	2554	0	+1	0	0	Сахалин

Примечание: знак «-» в таблице означает количество убывших, знак «+» – количество добавленных, «0» – отсутствие каких-либо изменений.

Выявленные группы разночтений могут быть как просто ошибками подготовки данных, так и отражать несколько разные подходы к агрегации данных, используемые для составления Общего свода и погубернских изданий.

Второй этап источниковедческого анализа агрегированных данных заключался в выявлении и корректировке ошибок в погубернских выпусках результатов Переписи 1897 г. (Первая Всеобщая перепись..., 1905) путем перекрестной проверки итоговых значений из первоисточника и значений, полученных путем агрегации данных, внесенных в базу данных «Профессии и занятия» (чисел по административным единицам и по группам занятий). Изначально такой прием использовался для проверки корректности внесения сведений в базу данных, но оказалось, что таким образом можно выявлять не только ошибки ввода, но и ошибки в печатных изданиях, например, такие как «смещение» строк, непечатанные или плохочитаемые символы, перестановка цифр местами (13 613 вместо 31 613), ошибки набора (2855 вместо 9855) и т.д. Всего в запрос попало 44 таких ошибки с общей разницей в 37 603 чел., что с учетом численности населения Сибири составило менее 1 %.

Наибольшее количество ошибок за счет «смещения» строк было выявлено по самозанятому населению в Баргузинском и Акшинском уездах Забайкальской области. Самые большие (по числу человек) ошибки были обнаружены для женщин – членов семей лиц, занятых в группе 17 «Сельское хозяйство», Приморской области (разночтение составило 18 тыс. чел.) (рис. 1), женщин – членов семей по Минусинскому уезду Енисейской губернии (10 тыс. чел.); мужчин, занятых в группе 19 «Животноводство», в Верхнеудинском округе Забайкальской области (разница 7 тыс. чел.), и мужчин, занятых в группе 22 «Рыболовство и охота», в уездах Приморской области (отличие в 3 тыс. чел.).

В отдельных случаях такие ошибки уже являются статистически значимыми, но не всегда могут быть очевидны как в громоздких таблицах источника, так и в многомерных отчетах базы данных или электронных таблицах информационной системы. Поэтому было принято решение, что в базе данных в качестве основных хранятся скорректированные значения, которые выделяются красным цветом в таблицах и отчетах, а значения из источника показаны во всплывающих окошках (см. рис. 1).

Первая Всеобщая перепись населения Российской империи 1897 г.
Распределение самозанятого населения по 65 группам занятий

Административная единица: Приморская область Население: Общее население городов, уездов, округов

Виды и группы видов деятельности	Занятое население			Члены семей	
	Все	Муж	Жен	Муж	Жен
Приморская область	113 657	108 888	4 769	43 173	66 506
11. Врачебная и санитарная деятельность	696	630	66	114	316
12. Служба при благотворительных учреждениях	4	4	0	1	4
13. Деятельность и служба частная. Прислуга, поденщики	15 126	13 507	1 619	856	1 428
14. Живущие доходом с капитала и недвижимого имущества или на средства, получаемые от родителей и родственников	889	512	377	550	639
15. Живущие на счет казны и общественных учрежд. и на средства частных лиц	981	671	310	30	79
16. Лишенные свободы и отбывающие наказание	939	931	8	5	4
17. Сельское хозяйство (Земледелие)	17 881	17 459	422	23 631	31 613
18. Пчеловодство и шелководство	36	35	1	2	4
19. Животноводство	4 347	4 099	248	5 129	8 756
20. Лесоводство и лесные промыслы	4 602	4 584	18	180	309
21. Рыболовство и охота	8 670	8 508	162	7 577	13 377
22. Добыча руд и копи	1 080	1 080	0	137	250
23. Выплавка металлов					

В источнике: 31613

Рис. 1. Скорректированные значения и данные источника в информационной системе «Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года»

Итак, если сопоставление разных изданий с результатами Переписи 1897 г. по численности населения Сибири и по группам занятий не дало статистически заметных различий, хоть и показало некоторые несоответствия в данных, то изучение ошибок источника показало в отдельных случаях наличие значимых расхождений. При этом количество ошибок и общая сумма

разночтений не влияют на показатели по Сибири или губерниям в целом. Таким образом, можно сделать вывод о высоком уровне подготовки статистических данных Переписи 1897 г.

Компаративный анализ опубликованных и персональных данных Переписи 1897 г. по численности и занятости населения г. Тобольска

Идея следующего этапа нашего исследования заключалась в сравнении агрегированных данных Переписи 1897 г. из базы данных «Профессии и занятия» и персональных сведений из переписных листов 1897 г., полученных из базы данных «Население городов» по населению города Тобольска по следующим характеристикам: численность, доля занятого населения и распределение его по группам занятий (на примере женщин).

Первым этапом работы стало определение и сопоставление численности Тобольска. В соответствии с опубликованными данными наличное население города составляло 20 425 чел. (Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года). База данных «Население городов» в настоящее время содержит сведения о 20 826 персонах. Стоит отметить, что в базу вносились все сведения из источника, даже записи, которые, очевидно, не учитывались в статистике наличного населения:

– вычеркнутые из переписных листов, наиболее часто такие записи встречались в Тюремном замке и Каторжной тюрьме (пересыльной);

– данные о «временно отсутствующих» в домохозяйствах, а также находящихся во «временной отлучке» или «командированных» в другие учреждения, города, уезды, регионы, например, солдаты Тобольского пехотного резервного батальона были записаны в казармах с пометкой о временной отлучке и проживании «на квартирах», в том числе и в качестве денщиков при офицерах.

Всего подобных записей было выявлено 179. Такая ситуация обусловлена технологией проведения переписи и заполнения переписных листов. В соответствии с Наставлением городским счетчикам счетчик заранее обходил вверенные ему домохозяйства и раздавал переписные листы для заполнения, а в день, к которому была приурочена перепись, «все переписные листы должны быть приведены к наличному составу населения» (Наставление городским счетчикам..., 1896, с. 4, 7). Указанные записи имеют соответствующую отметку в базе данных, поэтому в дальнейших подсчетах не учитывались.

Для выявления причин разночтений между базами данных был составлен запрос на полное соответствие фамилии, имени, отчества и возраста. В результате был получен список из 103 повторяющихся персон, не имеющих отметок об отсутствии или временном присутствии в домохозяйстве. Распространенной ситуацией была запись персоны в двух домохозяйствах: месте проживания и месте службы. При этом в ряде случаев не было очевидно, где именно человек проживал постоянно. Часто это были подростки (14–16 лет), записанные один раз в семье родителей, а другой: 1) в домохозяйстве, где они работали, например, прислугой; 2) учебных заведениях; 3) у других родственников (бабушки, дяди, брата) и т.п. Кроме того, некоторые записи не попали в результаты запроса в связи с техническими ошибками ввода, разным написанием личных имен (Мария / Марья, Никитишна / Никитична), изменением возраста на один год (10–11). Подобные записи не имеют никаких отметок в базе данных, поэтому не могут учитываться для дальнейших подсчетов.

В целом с учетом отметок («запись вычеркнута полностью», «временно отсутствует», «временно присутствует» и др.) в выборку по базе данных «Население городов» попали 20 627 чел., разночтения с агрегированными данными составили 1 %. Стоит отметить, что корректировка данных переписных листов (выявление повторов, отсутствующих и других погрешностей) была произведена губернским статистическим комитетом, т.е. счетчиками и заведующими счетных и переписных участков, «вручную» и на очень высоком уровне. Так, в вышедшем в конце 1897 г. Обзоре Тобольской губернии численность населения Тобольска, по предварительным данным переписи, составила 20 427 чел. (Статистический обзор..., 1898, с. 30).

Задачами второго этапа работы с базой данных «Население городов» было определение доли самозанятого населения и распределение его по группам занятий. Для примера мы выбрали женское население. Его численность по базе данных составила 9720 чел. (по агрегированным материалам – 9771 чел.), разнотчение составило 0,52 %. По агрегированным данным число самозанятых женщин составило 3331 чел. (или 34 % от женского населения). Всего были выявлены 44 группы, где были представлены женщины (Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года).

Несмотря на проведенную при создании базы данных «Население городов» унификацию занятий (например, запись переписного листа «состоит Тобольскою старшею городской оспеницею» формализовалась в занятие «врач, оспенник»), возможности базы данных не позволили «автоматически» рассчитать не только число самозанятого населения, но и распределить его по группам занятий. Предложенная разработчиками Переписи 1897 г. группировка данных по занятиям не раз подвергалась критике исследователями, которые считали ее неудобной из-за отсутствия обобщенных групп [Кадомцев, 1909, с. 19; Гозулов, 1970, с. 120; Миронов, 2000, с. 307]. Однако при попытке распределения записей из базы данных «Население городов» на категории «самозанятые / члены семей» и группы занятости также возникли сложности. Рассмотрим подробнее некоторые особенности сопоставления агрегированных групп со сведениями о занятиях из переписных листов в базе данных.

Одним из сложных моментов для формирования групп занятий стало включение в состав самозанятого населения учащихся, заключенных, обитателей сиропитательных и богоугодных заведений, поскольку, например, у учащихся графы о занятиях чаще всего оставались пустыми, для заключенных и больных указывалось прежнее занятие (до заключения или больницы) (рис. 2). Эта особенность требовала учета не только сведений граф о занятиях, но и данных из обложек об учреждениях (табл. 3, см. группы 15.1–15.4, 16.1).

≡ Население городов Сибири на рубеже XIX–XX вв.

Яркова Устинья Дмитриевна

г. Тобольск, Тобольский округ, Тобольская губерния
Площадь Соборная, дом Тюремный замок
Переписной участок № 1, Счетный участок № 1

Государственный архив г. Тобольска
ФОНД 417 ОПИСЬ 2 ДЕЛО 6 ПЕРЕПИСНОЙ № ЛИСТА 16 АРХИВНЫЙ № ЛИСТА 31

ВОЗРАСТ	31 год
СЕМЕЙНОЕ ПОЛОЖЕНИЕ	замужем
СОСЛОВИЕ	Крестьяне крестьянка из госуд.
ВЕРОИСПОВЕДАНИЕ	Православие
ЯЗЫК	Русский
ЗАНЯТИЕ ОСНОВНОЕ	торговец торгует на базаре рыбой
МЕСТО РОЖДЕНИЯ	Тобольский округ, Тобольская губерния Тобольской губернии и округа
МЕСТО ПРИПИСКИ	Бегишевская волость, Тобольский округ, Тобольская губерния Тобольской губернии и округа Бегишевской волости дер. Ярковой
МЕСТО ПРОЖИВАНИЯ	г. Тобольск, Тобольский округ, Тобольская губерния Тобольской губернии и округа (зачеркнуто) здесь

Рис. 2. Сведения о заключенной Тюремного замка в информационной системе «Население городов Сибири на рубеже XIX–XX вв.»⁴

Состав некоторых групп не всегда был очевидным. Например, из названия вида занятости 15.6 «Живущие пособием на местах жительства» (см. табл. 3) не понятно, что в него включали призреваемых и воспитанников, проживавших в семьях на средства Приказа общественного призрения.

Еще одной особенностью работы с базой данных является то, что для корректного подсчета самозанятого населения из сведений переписного листа в отдельных случаях необходимо было учитывать не только запись в графе «Основное или дополнительное занятие», но и записи в графе «Отношение в главе домохозяйства», поскольку в ней иногда содержались сведения, которые могли иметь отношение к занятости или источникам дохода человека. Например, могли быть сведения о воспитанниках и призреваемых, прислуге и работниках, при этом графы о занятии могли оставаться пустыми.

Стоит отметить, что в классификации занятий по Переписи 1897 г. присутствует учет одинаковых занятий в разных профессиональных сферах. Так, женщины, работающие прислужкой, при агрегации данных могли быть учтены в разных группы. Например, в группу 11 «Врачебная и санитарная деятельность» входили не только врачи, аптекари, сестры милосердия, но и прислуга в больницах (служанки, кухарки, прачки и т.п.) и аптеках (в Тобольске из 13 чел. – по агрегированным данным и 20 чел. – по переписным листам). Кроме того, прислуга учитывалась в группах занятых в учебных заведениях и богадельнях. Включение прислуги в разные группы создавало сложности подсчета в самой большой по численности занятых женщин группе 13 «Деятельность и служба частная. Прислуга, поденщики», которая включала по агрегированным данным 1178 чел., а по переписным листам – 1125 чел. Аналогичная проблема возникла с прачками, для которых был выделен вид занятости 62.3 «Прачешные и гладильни» (191 и 234 чел. соответственно), но они могли быть учтены и как прислуга в учреждениях образования или здравоохранения. Это, возможно, стало причиной разночтения в подсчетах в агрегированных данных и по сведениям переписных листов.

Еще одним примером неочевидного разделения самозанятых является группа 60 «Трактиры, гостиницы, меблированные комнаты, клубы». Она включала такие виды занятости, как «Содержание нахлебников и квартирантов», которая в таблице агрегированных данных по Тобольской губернии пропущена, и «Содержание гостиниц, меблированных комнат, ночлежных домов, постоянных дворов» со значением «0» (см. табл. 3, группы 60.1–60.4). Однако, судя по данным переписных листов, в Тобольске были женщины, которые и содержали постоянные дворы (5 чел.), и держали нахлебников и квартирантов (15 чел.).

Некоторые сложности вызывали малочисленные группы: «Телеграфисты, телеграфистки, телефонистки и почтальоны», «Фотография, фотографическое производство» и проч. Так, удалось установить телефонистку Ревердатто Екатерину Ивановну (28 лет), вдову французского подданного (родной язык – французский), и «хозяйку фотографии» Уссаковскую Марию Михайловну (26 лет), но выявить по базе данных «стипендиатку различных учреждений и частных лиц» не удалось (см. табл. 3, группа 15.5).

В целом количество самозанятых женщин по персональным данным составило 3534 чел. (34,5 % от женского населения Тобольска), разница с опубликованными результатами (3331 чел.) составила 6 %. Стоит отметить, что при подсчетах учитывалось только основное занятие, при наличии двух записей в графе (например, «при родителях» и «няня») мы относили персону к самозанятым и считали в соответствующей группе занятий.

Нам удалось распределить самозанятых по группам и некоторым видам занятий, а существенные разночтения были выявлены только для нескольких групп (например, прачки). Вместе с тем стоит отметить, что возможности баз данных не позволили автоматизировать сопоставление агрегированных и персональных данных, что обусловлено как сложной классификацией занятости (включающей группы и виды занятости с дублирующими профессиями) в опубликованных изданиях, пропусками данных в статистических материалах, так и особенностями работы с данными переписных листов, а именно необходимостью учета не только сведений графы об основном занятии, но и графы «Отношение к главе домохозяйства», а также места службы или содержания. В целом наш опыт по реконструкции методики агрегации сведений о занятиях в профессиональные группы Переписи 1897 г. подтверждает добросовестность работы статистических комитетов по проверке и подготовке данных о занятости, несмотря на громоздкую и сложную структуру профессиональных групп.

Таблица 3

Сравнение отдельных групп и видов занятий женского населения г. Тобольска по данным опубликованных источников и переписных листов⁵

Группы	Численность групп		Учреждения / комментарии
	По опубл. данным	По переписным листам	
Группы, для которых необходим учет данных об учреждениях (из обложек)			
15. Живущие на счет казны и общественных учреждений и на средства частных лиц	509	503	
15.1. Воспитанники закрытых учебных заведений, частных пансионатов и т.п.	288	289	Ученики и воспитанники Мариинской женской школы, Епархиального женского училища и Повивально-фельдшерской школы
15.2. Призреваемые в богадельнях, приютах и т.п.	105	103	Обитатели общественного призрения и мещанской богадельни, Александровского детского приюта
15.4. Больные в больницах	31	37	Больные Тобольской губернской больницы
15.5. Стипендиаты различных учреждений и частных лиц	1	Нет	Выявить не удалось
15.6. Живущие пособием на местах жительства	77	74	Призреваемые и воспитанники, проживающие в семьях на средства Приказа общественного призрения
16. Лишенные свободы и отбывающие наказание	23	21	
16.1 Арестанты в тюрьмах	23	21	Арестантки Тюремного замка
Пропущенные подгруппы в агрегированных данных			
60. Трактиры, гостиницы, меблированные комнаты, клубы	24	22	
60.1. Содержание гостиниц, меблированных комнат, ночлежных домов, постоянных дворов	0	5	
60.2. Содержание ресторанов, трактиров, буфетов, кухмистерских и чайных	6	2	
60.6. Содержание нахлебников и квартирантов	Нет	15	

Заключение

Проведенные эксперименты по источниковедческому анализу баз данных по агрегированным и персональным данным Первой всеобщей переписи населения позволяют сделать несколько выводов.

Во-первых, базы данных дают широкие возможности работы со структурированными данными, к которым относятся статистические материалы. Их инструменты позволяют провести как сопоставление разных изданий переписи, так и анализ одного выпуска, получить не только представление об особенностях структуры данных, но и выявить ошибки в опубликованных результатах Переписи 1897 г. При этом ограничения возможностей анализа в данном случае будут обусловлены именно структурой данных, заложенной в источнике.

Во-вторых, базы данных позволяют аккумулировать, унифицировать и хранить сведения массовых источников, к которым относятся переписные листы 1897 г., и формировать основу для их комплексного анализа, в том числе источниковедческого. В ходе нашего исследования изучение отметок переписного листа о временном отсутствии или присутствии, выбытии, вычеркнутых строк позволило понять технику подсчета населения в ходе проведения переписи. Работа с данными о занятости потребовала изучения источника с точки зрения методики внесения сведений в переписные листы и обложки, особенности записи разных видов и форм занятости, а также сведений о членах семей. В данном опыте ограничениями стали уровень унификации и структура самой базы данных по сведениям переписных листов, которая не предусматривала возможности автоматического сравнения персональных и агрегированных данных. Тем не менее сравнительный анализ был проведен и позволил осуществить реконструкцию методики агрегации сведений о занятости в профессиональные группы Переписи 1897 г.

В-третьих, источниковедческий анализ статистических материалов Переписи 1897 г., компаративный анализ баз данных по сведениям о численности населения города Тобольска и женской занятости позволяют сделать вывод о качественной проверке и подготовке материалов Переписи 1897 г. уже на уровне губернских статистических комитетов (разночтение по численности населения города по агрегированным и персональным данным составило 1 %).

Стоит отметить, что все используемые в статье агрегированные и первичные материалы Переписи 1897 г., а также базы данных доступны для пользователей онлайн. Таким образом, развитие цифровых ресурсов не только расширяет возможности исследователей, но и обеспечивает верификацию данных и выводов.

Примечания

¹ Исследование выполнено за счет гранта Российского научного фонда № 19-78-10020, <https://rscf.ru/project/19-78-10020/>.

² Источник: (Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года).

³ Источник: (Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года).

⁴ Источник: (Население городов Сибири на рубеже XIX–XX вв. по материалам Первой всеобщей переписи населения 1897 г.).

⁵ Источники: (Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года; Население городов Сибири на рубеже XIX–XX вв. по материалам Первой всеобщей переписи населения 1897 г.).

Список источников

Государственный архив в г. Тобольске (ГАТ). Ф. 417. Оп. 2. Д. 1–38.

Население городов Сибири на рубеже XIX–XX вв.: по материалам Первой всеобщей переписи населения Российской империи 1897 г.: база данных RU 2021620150 / *Е.А. Брюханова, Д.Н. Иванов, Н.В. Неженцева, О.И. Чекрыжова*; правообладатель Алт. гос. ун-т. № 2021620032; заявл. 13.01.2021; опублик. 21.01.2021. Бюл. № 2.

Население городов Сибири на рубеже XIX–XX вв. по материалам Первой всеобщей переписи населения 1897 г. [Электронный ресурс]. URL: <https://person1897.histcensus.asu.ru/> (дата обращения: 06.11.2023).

Наставление городским счетчикам. Первая всеобщая перепись населения Российской империи. СПб., 1896.

Общий свод по Империи результатов разработки данных первой всеобщей переписи населения, произведенной 28 января 1897 года. СПб., 1905. Т. II.

Первая Всеобщая перепись населения Российской империи 1897 г. СПб., 1905. Т. LXXII, LXXX. Профессии и занятия населения Российской империи конца XIX – начала XX веков (по материалам Всероссийской переписи населения 1897 года): база данных RU 2014621104 / *Е.А. Брюханова, Д.Н. Иванов*; правообладатель Алт. гос. ун-т. № 2014620481; заявл. 16.04.2014; опубл. 20.09.2014.

Профессии и занятия населения Российской империи конца XIX – начала XX века. Анализ данных Первой всеобщей переписи населения 1897 года [Электронный ресурс]. URL: <https://stat1897.histcensus.asu.ru/> (дата обращения: 06.11.2023).

Распределение населения по видам главных занятий и возрастным группам по отдельным территориальным районам. СПб., 1905. Т. 4.

Статистический обзор Тобольской губернии за 1897 г. Тобольск: Тип. губерн. управления, 1898.

Библиографический список

Акашева А.А. Книжные издания всероссийской переписи населения 1897 г. как первый опыт широкомасштабной публикации данных в государственной статистике. Источниковедческое изучение средствами культурной аналитики [Электронный ресурс] // *История*. 2023. Т. 14, вып. 6 (128). URL: <https://history.jes.su/s207987840027122-3-1/> (дата обращения: 06.11.2023).

Брюханова Е.А., Иванова Н.П. Документоведческий и источниковедческий анализ первичных материалов Первой всеобщей переписи населения Российской империи 1897 г. // *Вестник Том. гос. ун-та*. 2019. № 442. С. 96–107.

Бородкин Л.И. Историк и мир (больших) данных: вызовы цифрового поворота // *Историческая информатика*. 2019. № 3. С. 14–30.

Владимиров В.Н. Историческая информатика: пути развития // *Вестник Том. гос. пед. ун-та*. 2006. Вып. 1 (52). С. 86–92.

Володин А.Ю. Digital Humanities (цифровые гуманитарные науки): в поисках самоопределения // *Вестник Перм. ун-та. История*. 2014. № 3 (26). С. 5–12.

Гарскова И.М. Историческая информатика: эволюция междисциплинарного направления. СПб., 2018.

Гозулов А.И. Переписи населения земного шара. М., 1970.

Динамика экономического и социального развития России в XIX – начале XX вв. [Электронный ресурс]. URL: <https://www.hist.msu.ru/Dynamics/index.html> (дата обращения: 06.11.2023).

Исторические информационные системы: теория и практика / *С.И. Корниенко* [и др.]. М., 2021.

Кадошцев Б.П. Профессиональный и социальный состав населения Европейской России по данным переписи 1897 г. СПб., 1909.

Корнев М.С. Источниковедение 2.0: о новых подходах к работе с источниками в сетевой цифровой среде // *Вестник РГГУ. Литературоведение. Языкознание. Культурология*. 2018. № 11. С. 59–66.

Миронов Б.Н. Социальная история России периода империи (XVIII – начало XX в.): в 2 т. Т. 1. Генезис личности, демократической семьи, гражданского общества и правового государства. СПб., 2000.

Можжаева Г.В. Digital Humanities: цифровой поворот в гуманитарных науках // *Гуманитарная информатика*. 2015. Вып. 9. С. 8–23.

Ристат: электронный архив Российской исторической статистики [Электронный ресурс]. URL: <https://ristat.org/> (дата обращения: 06.11.2023).

Юмашева Ю.Ю. Документные ресурсы архивов, библиотек и музеев в сети Интернет [Электронный ресурс] // *Историческая информатика*. 2018. № 1. С. 1–13. URL: https://nbpublish.com/library_read_article.php?id=25513 (дата обращения: 06.11.2023).

Юмашева Ю.Ю. Актуальные проблемы источниковедения информационной эпохи [Электронный ресурс] // *Документ. Архив. История. Современность*. Екатеринбург, 2023. Вып. 23. С. 109–140. URL: <http://elar.urfu.ru/handle/10995/126651> (дата обращения: 06.11.2023).

Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition online. Available at: <https://hsus.cambridge.org/HSUSWeb/HSUSEntryServlet> (accessed: 06.11.2023).

Mining Microdata: Economic Opportunity and Spatial Mobility in Britain and the United States, 1850–1881 / K. Inwood [et al.] // IEEE International Conference on Big Data (Big Data) 2014. Available at: <https://ieeexplore.ieee.org/document/7004446> (accessed: 06.11.2023).

Ruggles S. Big Microdata for Population Research // Demography. 2014. No. 51(1). P. 287–297.

Дата поступления рукописи в редакцию 21.11.2023

DATABASES AND SOURCE ANALYSIS IN THE DIGITAL AGE: OPPORTUNITIES AND LIMITATIONS (THE EXAMPLE OF SIBERIA IN THE 1897 CENSUS DATA)

E. A. Bryukhanova

Altai State University, Lenin av., 61, 656049, Barnaul, Russia

elena@hist.asu.ru

ResearcherID: N-9527-2016

Scopus Author ID: 57194780366

SPIN: 1342-7026

N. P. Ivanova

Altai State University, Lenin av., 61, 656049, Barnaul, Russia

Ivanovanp@gmail.com

ResearcherID: JPK-0593-2023

SPIN: 5829-2234

The modern stage of historical and social sciences development is characterized by an active creation of digital projects and resources. Within these projects, new types of sources are being formed, such as digitized documents, databases, and information systems. These trends expand the possibilities for researchers and set new challenges for them, as well as open new scientific directions. One such direction is the source study of information resources and digital data, which is particularly relevant for historical sources that have a debatable nature, such as the materials of the 1897 First General Census of the Russian Empire. The aim of this study is to assess the possibilities of database technology in analyzing the 1897 Census materials, using the data on the number and employment of Siberia as an example. In the study, two interrelated tasks were addressed: to consider the advantages and limitations of using databases for the analysis of the source materials from the 1897 Census, and to assess the reliability of census data for particular indicators. The sources of this article include both published and archival materials of the 1897 Census (aggregated data and personal information of Tobolsk census forms), as well as databases created from these materials. The study was based on a systematic approach and utilized retrospective, comparative methods, as well as internal criticism of sources for source analysis. The authors state that databases have significant potential for source analysis and that the collection and processing of the 1897 Census materials for the Siberian population was of high quality: the discrepancy in the number of Tobolsk according to aggregated and personal data was only 1%. In addition, the development of digital resources and access to the 1897 Census materials and databases for users not only expands the capabilities of researchers, but also ensures the verification of data and conclusions.

Key words: source analysis, statistical materials, census forms, database, 1897 Census, female employment, Tobolsk, Siberia.

Acknowledgments

¹ The reported study was funded by the Russian Science Foundation, project № 19-78-10020 “Urbanization processes in Siberia at the turn of the 19th–20th centuries: a complex approach to urban population analysis based on the materials of 1897 Russian Empire General Population Census”, <https://rscf.ru/project/19-78-10020/>.

References

Akashcheva, A.A. (2023), “Book editions of the All-Russian Population Census of 1897 as the first experience of large-scale publication of data in state statistics. Source study by means of cultural analytics”, *Istoriya*, vol. 14, issue 6 (128), available at: <https://history.jes.su/s207987840027122-3-1/> (accessed 06.11.2023).

Baskerville, P., Dillon, L., Inwood, K., Roberts, E., Ruggles, St., Schürer, K. & R. J. Warren (2014), “Mining microdata: Economic opportunity and spatial mobility in Britain and the United States, 1850–1881”, in *IEEE International Conference on Big Data (Big Data) 2014*, available at: <https://ieeexplore.ieee.org/document/7004446> (accessed 06.11.2023).

- Borodkin, L.I. (2019), “Historian and the world of (big) data: challenges of digital turn”, *Istoricheskaya informatika*, № 3, pp. 14–30.
- Bryukhanova, E.A. & N.P. Ivanova (2019), “Document study analysis and source study analysis of the first 1897 General Census of the Russian Empire: nominative data”, *Vestnik Tomskogo gosudarstvennogo universiteta*, № 442, pp. 96–107.
- Dinamika ekonomicheskogo i sotsial'nogo razvitiya Rossii v XIX – nachale XX vv.* [Dynamics of economic and social development of Russia in the 19th – early 20th centuries] (n.d.), available at: <https://www.hist.msu.ru/Dynamics/index.html> (accessed 06.11.2023).
- Garskova, I.M. (2018), *Istoricheskaya informatika: evolyutsiya mezhdistitsiplinarnogo napravleniya* [Historical computer science: the evolution of the interdisciplinary field], Aleteya, St. Petersburg, Russia, 408 p.
- Gozulov, A.I. (1970), *Perepisi naseleniya zemnogo shara* [World censuses], Statistika, Moscow, USSR, 176 p.
- Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition online* (n.d.), available at: <https://hsus.cambridge.org/HSUSWeb/HSUSEntryServlet> (accessed 06.11.2023).
- Kadomtsev, B.P. (1909), *Professional'nyy i sotsial'nyy sostav naseleniya Evropeyskoy Rossii po dannym perepisi 1897 g.* [Occupational and social composition of the European Russian population according to the 1897 census], Izdanie S.-Peterburgskogo Politekhnicheskogo instituta, St. Petersburg, Russia, 110 p.
- Kornev, M.S. (2018), “Source studies 2.0. New approaches to working with sources in a network digital environment”, *Vestnik RGGU. Seriya «Literaturovedenie. Yazykoznanie. Kul'turologiya»*, № 11, pp. 59–66.
- Kornienko, S.I., Gagarina, D.A. & N.G. Povroznik (2021), *Istoricheskie informatsionnye sistemy: teoriya i praktika* [Historical information systems: theory and practice], Izd. dom Vysshey shkoly ekonomiki, Moscow, Russia, 231 p.
- Mironov, B.N. (2000), *Sotsial'naya istoriya Rossii perioda imperii (XVIII – nachalo XX v.). T. 1. Genezis lichnosti, demokraticheskoy semyi, grazhdanskogo obshchestva i pravovogo gosudarstva* [Social history of Russian Empire (18th – early 20th century), vol. 1. Genesis of individual, democratic family, civil society and the constitutional state], Dm. Bulanin, St. Petersburg, Russia, 2000, 583 p.
- Mozhaeva, G.V. (2015), “Digital Humanities: Digital Turn in the Humanities”, in *Gumanitarnaya informatika* [Humanities Computer Science], issue 9, Izdatel'stvo Tomskogo gosudarstvennogo universiteta, Tomsk, Russia, pp. 8–23.
- Ristat: Elektronnyy arkhiv Rossiyskoy istoricheskoy statistiki* [Ristat: Electronic Repository of Russian Historical Statistics] (n.d.), available at: <https://ristat.org> (accessed 06.11.2023).
- Ruggles, St. (2014), “Big Microdata for Population Research”, *Demography*, № 51(1), pp. 287–297.
- Vladimirov, V.N. (2006), “Historical Computer Science: ways of development”, *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*, № 1 (52), pp. 86–92.
- Volodin, A.Yu. (2014), “Digital Humanities in Search of Self-Defining”, *Vestnik Permskogo universiteta. Seriya «Istoriya»*, № 3 (26), pp. 5–12.
- Yumasheva, Yu.Yu. (2018), “Document Resources of Archives, Libraries and Museums in the Internet”, *Istoricheskaya informatika*, № 1, pp. 1–13, available at: https://nbpublish.com/library_read_article.php?id=25513 (accessed 06.11.2023).
- Yumasheva, Yu.Yu. (2023), “Contemporary Challenges in Source Studies of the Information Age”, in *Dokument. Arkhiv. Istoriya. Sovremennost'* [Document. Archive. History. Modernity], vol. 23, Izdatel'stvo Ural'skogo universiteta, Yekaterinburg, Russia, pp. 109–140, available at: <http://elar.urfu.ru/handle/10995/126651> (accessed 06.11.2023).