

Сведения о выполненных работах и полученных научных результатах
в период с 14.08.2023 г. по 30.06.2024 г.

по проекту **«Исторические информационные ресурсы в исследовательской
инфраструктуре: модели репрезентации, анализа и интеграции данных
по региональной истории»**,
поддержанному Российским научным фондом

Соглашение № 23-78-10119

Руководитель: канд. ист. наук Полежаева Татьяна Владимировна

В результате реализации первого этапа проекта была разработана аналитическая база данных «Библиотеки об истории Сибири», включившая 314 информационных ресурсов 189 центральных библиотек регионов и вузов Сибирского, Дальневосточного и Уральского федеральных округов. Указанные ресурсы были описаны по 35 параметрам. Созданный аналитический корпус включил разнородные источники по тематике (история организации, история региона, история страны), типу ресурса (полнотекстовые электронные библиотеки / коллекции, краеведческие порталы или веб-страницы, виртуальные выставки, электронные (био)-библиографические справочники), ценности (коллекции редких книг и книжных памятников).

Было установлено, что среди указанных информационных ресурсов по объему данных преобладают оцифрованные источники, чаще всего размещенные в электронных библиотеках с использованием специального программного обеспечения (самые популярные программы – ИРБИС, DSpace).

Была описана модель репрезентации регионального исторического наследия библиотеками, в которой сохраняются принципы и практики работы с физическими объектами, перенесенные в цифровую среду. Зафиксированы «болевые точки» для историков в области провенанса цифровых коллекций, репликации объектов, принципов размещения объектов в коллекции и инструментов работы с ними. В качестве перспективной модели интеграции исторических информационных ресурсов библиотек в научные коммуникации предложено использование принципов «коллекции как данные» и происхождения данных (data lineage). Последний усложняет традиционную хронологию, фиксируя все изменения объекта в цифровой среде и рассматривая движение информации от источника по точкам обработки и применения.

В качестве базового критерия определения востребованности ресурсов, создаваемых библиотеками, было определено цитирование коллекций. Оно определялось по ссылочному аппарату публикаций, проиндексированных в РИНЦ. Установлено, что лидерами по цитированию являются электронные библиотеки, реализованные на DSpace. Выделена характерная особенность исследовательского информационного поведения: предпочитают цитировать бумажную версию, даже если работали с электронной. Кроме того, в качестве косвенного критерия востребованности использовалось количество благодарностей

библиотекарям, которые присутствуют в соответствующих разделах публикаций. Так, по базе Web of Science было подсчитано, что за 2019-2023 гг. российские историки поблагодарили библиотеки 356 раз, из них сибирские библиотеки получили 3 благодарности.

В ходе реализации проекта была разработана карта рисков и угроз сохранности исторической информации в цифровой среде. Для оцифрованных ресурсов были выделены риски, возникающие: а) на этапе отбора и подготовки объекта для оцифровки, б) на этапе машинной обработки и АТР (automatic text recognition), в) публикации, обеспечения доступа и использования данных. Для born-digital материалов актуальна группа рисков в). Для текущей модели цифрового ресурса библиотек самой актуальной проблемой оказывается технологическая уязвимость. На ее преодоление могут быть направлены усилия государственной платформы Гостех (модуль «Наука» сейчас находится в разработке). Перспективными представляются разработки Solid(social linked data) Тима Бернерс Ли и проект пост-платформ А. Турского. Отдельно были описаны риски, связанные с внедрением искусственного интеллекта в технологические процессы библиотек.

Абрис текущей модели информационных ресурсов библиотек позволил перейти к проектированию научно обоснованной модели информационно-аналитического ресурса «Православный ландшафт таежной Сибири: акторы, институты, сети». На первом этапе проекта был сформирован перечень и дана характеристика методологических и методических подходов к представлению источников тематически связанных с конфессиональной (православной) историей региона, в том числе: а) обоснован подход к изучению и цифровой презентации религиозного ландшафта региона как системы официальных и неформальных социальных сетей, конкуренция которых продуцирует разные информационные потоки, формирует разные образы «православной Сибири» и разные типы стейкхолдеров, заинтересованных в их трансляции; б) определены параметры и содержание базовой, дополняющей и генерируемой информации онлайн ресурса с учетом задач развития «цифрового источниковедения» и использования born-digital материалов для расширения источниковой базы по религиозной истории региона и изучения процессов формирования групповых идентичностей и солидарностей (территориальных, религиозных, этнических и пр.).

Введена типология религиозных ландшафтов (древовидный и ризомный), которая в перспективе позволит визуализировать и учитывать основные конституирующие факторы православного ландшафта таежной Сибири – церковно-государственную программу строительства приходской жизни, интеллектуальную и хозяйственную активность старообрядческого населения.

Разработаны алгоритмы сбора конкретно-исторических данных и их хранения в цифровом формате, нацеленные: а) на раскрытие связей между разными социальными акторами и институтами, качеством приходской жизни и популярностью официальных и неофициальных культовых мест, ареалами распространения канонических и неканонических религиозных практик (шаблон сбора данных

включает 31 тематическую группу с 335 привязанными к географическим координатам дескрипторами); б) на обеспечение безопасного хранения и расширения цифровых данных, их анализа историческими и математическими методами.

По материалам епархиальной печати и клировых ведомостей приходов Томского уезда составлена база данных (в настоящее время – 3690 записей). Уточнены принципы и алгоритмы соединения и взаимной проверки нарративных и статистических источников при формировании фактографической базы онлайн ресурса. Использование типологии исторических данных (повторяющаяся, верифицируемая, уникальная информация) позволило определить региональные составляющие церковно-государственной презентации образа «православной Сибири» на рубеже XIX–XX вв., выявить новые корреляции между редакторской политикой и объемами предоставляемой информации о приходской жизни. С помощью математического анализа клировых ведомостей обнаружены разрывы и противоречия в создаваемой «сверху» и стихийно возникающей «снизу» сети религиозных сообществ региона (в числе новых данных – установленные зависимости между возрастом, образованием, опытом священников и темпами школьного строительства в приходах с большим числом староверов). Материалы будут использованы для комплексного представления половозрастного и этноконфессионального состава старожилов и переселенцев, проверки математическими методами их потенциальной готовности связывать свою религиозную жизнь с конкретными культовыми местами и конфессиональными лидерами.

В ходе реализации проекта было опубликовано 3 научных статьи, принято участие в 4 научных конференциях.