

УДК 024.5:004

ББК 78.375.0с51

DOI 10.25281/0869-608X-2024-73-2-129-142

Е.Д. ЖАБКО

Цифровые коллекции как данные: новое направление в работе библиотек

Реферат. Целью статьи является анализ теоретических основ и опыта деятельности зарубежных библиотечных учреждений по направлению «цифровые коллекции как данные». Цифровые коллекции рассматриваются не только как новые формы организации знаний, но как инструмент проведения научных исследований в области гуманитарных дисциплин. Практика показала, что огромные массивы оцифрованных текстов и изображений могут использоваться более успешно при условии применения современных информационных технологий. Объединение методологии исследований в гуманитарных и социальных науках и компьютерных средств позволяет извлекать количественные и географические данные, факты, изображения на основе автоматического анализа распознанных текстов, визуализировать и отображать данные в различных форматах, предоставлять развернутые метаданные, моделировать исторические процессы и многое другое. Приведены наиболее показательные примеры инновационных проектов в ряде европейских библиотек. Дается краткая характеристика современного состояния преобразования цифровых коллекций как данных в США. Определено, что в последние десятилетия в России сформировалась научная, образовательная и технологическая среда для исследований на базе ресурсов библиотек. Библиотеки должны преодолеть свою оторванность от научных исследований в сфере цифровых гуманитарных наук и войти в них в качестве равноправного партнера. Определены задачи отечественных библиотек по дальнейшему развитию цифрового контента с учетом достижений в области информационных технологий и современного уровня технических знаний.

Ключевые слова: библиотеки, архивы, музеи, цифровые коллекции как данные, лаборатории цифровых гуманитарных наук, цифровые коллекции, информационные технологии, открытые данные, зарубежный опыт, библиотечно-информационное обслуживание, национальные библиотеки, библиотеки образовательных учреждений.

Для цитирования: Жабко Е.Д. Цифровые коллекции как данные: новое направление в работе библиотек // Библиотековедение. 2024. Т. 73, № 2. С. 129–142. DOI: 10.25281/0869-608X-2024-73-2-129-142.



Елена Дмитриевна Жабко,
Информационный
историко-научный центр —
Военная историческая
библиотека
Генерального штаба
Вооруженных сил
Российской Федерации,
старший научный сотрудник
Дворцовая пл., д. 10,
Санкт-Петербург, 191055,
Россия

доктор педагогических наук
ORCID 0000-0002-0871-5874;
SPIN 9942-9780
edzhabko@gmail.com

Целью статьи является анализ теоретических основ и опыта деятельности зарубежных библиотечных учреждений по направлению «цифровые коллекции как данные» (Digital Collections as Data). «...В современных условиях цифровые коллекции воспринимаются не столько как информационный и маркетинговый буклет физической коллекции, сколько как публикационная платформа, задача которой — создание новых знаний и распространение знаний» [1, с. 58]. Источниковой базой стали монографии, публикации зарубежных и отечественных специалистов в периодических изданиях, сайты и порталы национальных библиотек (НБ) и архивов, профессиональных ассоциаций, отдельных глобальных и национальных проектов. Также использовались профессиональные блоги, являющиеся не только средством коммуникаций, но и источниками неопубликованной, но ценной информации по актуальным проблемам создания и эксплуатации различных цифровых продуктов и сервисов.

Российские специалисты, обращаясь к теме участия зарубежных библиотек в проектах цифровых гуманитарных наук, подробно охарактеризовали новые практики и существующие модели цифрового обслуживания [2], представили обзор методов искусственного интеллекта в исторической науке и прикладной деятельности библиотек, музеев и архивов [3], а также в весьма краткой форме дали характеристику направления [4].

В настоящей статье будут рассмотрены возможности использования коллекций оцифрованных текстов и изображений в библиотеках на основе применения современных информационных технологий, в том числе в рамках цифровых гуманитарных наук.

Новое направление работы с цифровым контентом за рубежом

Направление «коллекции как данные» появилось за рубежом в начале 2000-х гг., однако его активное развитие относится к 2013–2023 годам. Терминологически «цифровые коллекции как данные» определяются по-разному. В широком смысле это парадигма, направленная на развитие расширенного набора исследовательских, педагогических и художественных возможностей, основанных на компьютерном использовании коллекций культурного наследия [5]. В более практическом смысле «коллекции как данные» — это

новое направление деятельности, выход за рамки традиционной работы с коллекциями, извлечение данных для проведения гуманитарных исследований при помощи современных технологий [6]. Результатом этой деятельности является повышение исследовательской ценности библиотечных цифровых коллекций (не только библиотечных), переход от процессов формирования электронных библиотек к предоставлению наборов открытых данных. Одним из актуальных можно считать определение, приведенное в глоссарии итогового отчета проекта «Цифровые коллекции как данные: от части к целому», которое подчеркивает многоаспектность рассматриваемого явления: термин «цифровые коллекции как данные» относится к концепции, сообществу и области практики, фокусирующихся на разработке и использовании коллекций в качестве данных при помощи компьютерных технологий [7].

Как уже отмечалось, с 2013 г. в ряде европейских стран и в США стали активно реализовываться проекты, направленные на переосмысление места и роли цифровых коллекций в информационной исследовательской среде и на повышение эффективности их использования. Это предполагало поиск путей дальнейшего их развития. Коллекции, сформированные в библиотеках, архивах и музеях, в так называемой среде GLAM (Galleries, Libraries, Archives, Museums), представляли собой большие по объему массивы оцифрованных текстов и графических изображений. Изучение подобных ресурсов и извлечение необходимых количественных и географических данных, фактов, изображений требует не только времени исследователей, но специальных программно-технологических инструментов, не говоря уже про их дальнейшую обработку.

Выявление закономерностей и тенденций выполняет программное обеспечение, масштабно анализируя данные (обычный человек в короткий срок это сделать не может). Благодаря современным технологиям эти ресурсы, уже однажды созданные, могут быть повторно использованы в новых контекстах (re-use). Эта философия развивается в музейной сфере, где используется термин «культурные коллекции как данные» (cultural collections as data) [8], а также в сфере архивов, включая веб-архивы, — «архивы как данные» (archive as data) [9]. Общие подходы по предоставлению данных цифровых коллекций в сетевом пространстве не ограничиваются только органи-

зационными и технологическими вопросами (содержание данных, целевая аудитория, направления использования, форматы, программное обеспечение), они подразумевают учет требований международного и национального законодательства, прежде всего авторского.

Развитие этого направления в библиотеках напрямую связано с созданием на их базе специализированных лабораторий [10; 11]. Первоначально они появились в американских университетах (например, в Гарвардском университете), крупных публичных библиотеках (Нью-Йоркская публичная библиотека, Библиотека Конгресса США), а затем и в Европе, в частности в Британской библиотеке и НБ Нидерландов. В ряде случаев подобные лаборатории, работающие на стыке библиотечных и компьютерных наук, создавались на основе сотрудничества различных учреждений, например Лаборатории цифровых исследований, созданной Королевской библиотекой Бельгии, и Центра цифровых гуманитарных наук Гентского университета¹. Эти лаборатории не обязательно находятся в структуре какого-либо учреждения, они могут быть виртуальными, а участники взаимодействуют в онлайн-режиме.

Тенденция создания лабораторий была характерна для всех типов библиотек — национальных, университетских, публичных. По сути, речь идет о библиотечных центрах цифровых гуманитарных наук, целесообразность базирования которых в библиотеках обусловлена наличием электронных фондов (цифровых коллекций), сформированных как в результате оцифровки традиционных изданий, так и объединения исходно электронных документов. Обязанности таких центров подробно охарактеризованы: формирование цифровых коллекций для научных или образовательных целей; создание средств идентификации, построения баз данных и их анализа; управление исследовательским процессом; поддержка веб-сайтов; цифровая публикация и хранение результатов; организация обучения в сфере цифровых гуманитарных наук, создание лекционных курсов и программ, организация семинаров и конференций, сотрудничество с аналогичными структурами [12].

Междисциплинарный характер нового направления деятельности нашел подтверждение в создании в 2018 г. международного сообщества GLAM Labs. Сообщество объединило учреждения нескольких десятков стран, в том

числе библиотеки. В 2019 г. была подготовлена и представлена в открытом доступе книга «Откройте лабораторию GLAM» [13], в которой описаны возможные модели организации лабораторий, а также определены их сущностные функции.

Цифровые коллекции в европейских учреждениях культуры

В рамках настоящей статьи особый интерес представляет опыт библиотек национального уровня. Так, НБ Шотландии реализует проекты, целью которых является применение вычислительных методов при работе с коллекциями. Это позволяет использовать все преимущества цифровых коллекций, а для библиотеки — стать местом поддержки цифровых наук и обеспечить «прозрачность» процессов формирования данных [14]. Пользователи могут применять методы извлечения данных из текстов, анализировать их в целях своих научных разработок. С 2019 г. на платформе Data Foundry коллекции стали предоставляться не только в виде оцифрованных материалов, но и коллекций метаданных, геопространственных и организационных данных, а в будущем планируется предоставление пользователям аудиовизуальных данных и данных веб-архива.

Основными принципами представления коллекций НБ Шотландии на платформе Data Foundry стали открытость, прозрачность и практичность. Открытость означает, что опубликованные данные доступны открыто и в форматах, пригодных для повторного использования. Под практичностью понимается предоставление наборов данных в дополнительных форматах, что обеспечивает их максимальную доступность. И, наконец, прозрачность — ответственное отношение к происхождению данных и информирование об источниках их получения [15]. Значимым проектом НБ Шотландии в рамках направления «коллекции как данные» в 2023—2024 гг. стало подробное исследование и масштабный анализ содержания 100 томов энциклопедии «Британника» (1768—1860). Работа по превращению коллекций в данные идет в НБ Шотландии в рамках программы массовой оцифровки фондов для достижения поставленной цели — оцифровать треть своих коллекций к 2025 году. Для этого потребовалось пересмотреть весь рабочий процесс оцифровки. Вместо того чтобы оцифровывать

документы исключительно для онлайн-галерей изображений, пришлось уделять больше времени предоставлению дополнительных форматов файлов и генерировать контекстную информацию о коллекциях.

Не осталась в стороне и Британская библиотека, которая в 2013 г. открыла специальную лабораторию для экспериментального использования своих цифровых коллекций. Интересно, что некоторые разработки лаборатории сразу были направлены на повышение эффективности сугубо внутрибиблиотечной деятельности, например на проверку обоснованности выбора документов для оцифровки. Известно, что в целом ряде случаев процесс отбора определяется практическими моментами: степенью сохранности книжных изданий, доступностью издания, размером книги и ее оформлением, наличием авторских прав, вкусовыми предпочтениями специалистов и др. Обычно это приводит к созданию нерепрезентативной оцифрованной коллекции, т. е. отличающейся от физической коллекции по ряду параметров. В связи с этим в 2013 г. Британской библиотекой был проведен анализ соотношения количественных характеристик цифровых коллекций и представленной в них тематики относительно тех традиционных фондов, которые подлежали переводу в электронную форму. Для этого была разработана программа поиска в массиве 1,9 млн записей книг XIX века. На момент проведения эксперимента только 2,6% книг были оцифрованы. Поисковый механизм позволил сделать репрезентативные выборки книг на основе отфильтрованных поисковых запросов как цифровых, так и физических объектов из более крупного корпуса для дальнейшего анализа [16]. Это дало специалистам Британской библиотеки понимание соотношения оцифрованных материалов, хранящихся в Британской библиотеке, относительно физических коллекций как по хронологическим характеристикам, так и по темам.

В 2013 г. Британская библиотека реализовала проект Mechanical Curator Collection, загрузив в фотохостинг Flickr Commons более 1 млн изображений (автоматически извлеченных из распознанных и свободных от авторских прав книжных изданий): иллюстраций, карикатур, рисунков, карт, ранних фотографий, копий настенных фресок, рождественских этикеток и пр. [17]. Источниками получения изображений стали 65 тыс. книг XVII–XIX веков. Все издания были оцифрованы корпорацией Microsoft и переданы Британской библиотеке для опубли-

кования без всяких ограничений. Изображения не только доступны для пользователей в целях ознакомления, но и сопровождаются информацией, описывающей историю происхождения и раскрывающей их содержание.

В 2015 г. Британская библиотека представила стратегию развития до 2023 г. под названием «Живое знание» (Living Knowledge: The British Library 2015–2023)², в которой в целом определялись приоритеты комплектования фондов и организации физического и интеллектуального доступа к создаваемым ресурсам. В 2017 г. Британская библиотека выработала новую стратегию исследования данных [18], а чуть ранее спроектировала виртуальное пространство для исследователей, позволяющее загружать значительные по объему массивы общедоступных данных и цифровых коллекций библиотеки. Таким образом, пользователи смогли изучать их, экспериментировать и разрабатывать собственные научные проекты.

Помимо этого, в Британской библиотеке было реализовано несколько десятков проектов, позволяющих работать с данными коллекций на основе полнотекстового поиска и анализа, преимущественно с коллекциями печатных изданий на западных языках и коллекциями газетных изданий. Для расширения возможностей информационных технологий Британская библиотека стала сотрудничать с Институтом Алана Тьюринга, а также рядом университетов. Благодаря этому сотрудничеству были внедрены новые методы искусственного интеллекта, а также проведен масштабный анализ множества материалов. Примером может служить совместный проект Британской библиотеки, Института Алана Тьюринга, Кембриджского университета, университетов Восточной Англии, Эксетера и Королевского колледжа Living with Machines³, посвященный влиянию промышленной революции на жизнь людей.

На конец 2023 г. в цифровом репозитории Британской библиотеки находится более 154 наборов разнообразных данных о цифровых коллекциях (оцифрованные еврейские рукописи, книжные издания XVII–XIX вв., театральные программки, карточный каталог отдела Азии и Африки и др.)⁴. Каждый набор данных имеет уникальный идентификатор цифрового объекта (DOI), на который пользователь может сослаться в своей работе. В настоящее время Британская библиотека публикует данные своих коллекций в зависимости от типа/формата данных, не только на British

Library Research Repository, но и на сторонних платформах, включая Europeana, фотохостинг Flickr, Викимедиа, Hugging Face и др. В ежегодном отчете Британской библиотеки за 2022—2023 гг. [19] указывается, что библиотека и далее будет работать в направлении использования искусственного интеллекта для обеспечения доступа к своим коллекциям, в том числе благодаря тому, что она возглавляет международную сеть «Искусственный интеллект для библиотек, архивов и музеев».

Процессы работы с оцифрованными массовыми документами с применением информационных технологий в Великобритании распространены и в архивной сфере. В рамках настоящей работы приведем только один пример. Национальный архив Великобритании еще в 2010 г. создал лабораторию по изучению возможности использования информационных технологий для обеспечения эффективного доступа к архивным фондам. Самым первым проектом стал «Поиск фотографий по истории Великобритании» (UK History Photo Finder), затем последовало несколько других значимых проектов. С 2014 г. Национальный архив Великобритании оцифровывал написанные от руки военные дневники времен Первой мировой войны, документирующие историю британской армии и ее подразделений на Западном фронте. Используя краудсорсинговую платформу Operation War Diary, историки проиндексировали 1,5 млн отсканированных и заархивированных страниц. Они классифицировали и комментировали важную информацию каждой страницы дневника, генерируя обширные временные ряды данных о воинских частях, включая информацию о потерях, численности подразделений, повседневной армейской жизни, военной деятельности, именах и званиях солдат, погоде, названиях мест дислокации и датах [20]. Это только один из имеющихся проектов. Как уже отмечалось, в ряде случаев тема «архивные цифровые коллекции как данные» рассматривается как самостоятельная. Также ведутся работы по проектам предоставления пользователям музейных данных.

Вернемся к национальным библиотекам. НБ Нидерландов, оцифровав свои газетные и журнальные фонды, на платформе Delpher обеспечивает доступ к данным 400 тыс. журналов XIX—XX вв. [21]. Адресные книги, альманахи, спортивные, ранние научные, общественно значимые, молодежные журналы и многое другое представлено для широкого доступа в виде отска-

нированных печатных страниц с распознаванием текста. Пользователям также доступны описательные и структурные метаданные. Описательные метаданные содержат библиографические сведения (автор, название или дата издания), структурные — предоставляют информацию о структуре файла (количество страниц, область ввода и абзацы). Это же касается коллекции из 2 млн голландских газет с 1618 по 1995 год. Этот ресурс является востребованным, поскольку газетные издания с такой глубокой ретроспективой являются значимым источником получения данных исторического характера.

НБ Финляндии сформировала коллекцию газет, издававшихся на территории Финляндии с 1771 по 1929 г. на шведском, финском и русском языках. Коллекция открыта для всех через сервис поиска по цифровым материалам⁵. Пользователям доступен газетный контент в виде статичных изображений в формате PDF и в текстовом формате (OCR). Библиотека в онлайн-режиме предоставляет не только инструменты для поиска, сбора и анализа газетных данных, но и некоторые расширенные функциональные возможности [22; 23].

Среди европейских библиотек нельзя не отметить эксперименты с коллекциями как данными в Королевской библиотеке Бельгии, где формируются устойчивые рабочие среды для облегчения создания корпуса в цифровых гуманитарных науках [24]. В 2020 г. библиотека приступила к реализации проекта DATA-KBR-VE 47 (2020—2024), целью которого является оптимизация информационно-коммуникационной инфраструктуры для стимулирования устойчивого доступа на уровне данных к оцифрованным и неоцифрованным коллекциям для гуманитарных исследований. Ключевым результатом проекта является разработка и внедрение платформы открытых данных для публикации коллекций Королевской библиотеки Бельгии в виде наборов данных. В рассматриваемом направлении работают также НБ Франции и Германии.

Следует отметить, что в ряде европейских стран реализуются проекты, которые не обязательно базируются в библиотеках, но используют в том числе библиотечный контент. В качестве примера можно привести цифровую платформу Национального центра документации Греции⁶, созданную в 2015 году. В настоящее время она предоставляет пользователям средства поиска, фильтрации, просмотра и визуального представления цифрового контента.

Важными для библиотек являются программы, проекты, мероприятия научного, методического, обучающего характера, которые проводятся в рамках европейской цифровой библиотеки EuropeanaPro. В документе «Стратегия Европеаны 2020–2025. Расширяя возможности цифрового шанса» [25] подтверждены уже определенные ранее приоритеты цифровой трансформации сферы европейского культурного наследия. В качестве основной из целей определена дальнейшая поддержка работы с информационными технологиями для открытия цифровых коллекций новыми способами. В конце 2023 г. был опубликован документ «Памятка по публикации коллекций как данных в учреждениях GLAM» [26], в котором даются рекомендации по всем направлениям деятельности учреждений в этой области, включая использование искусственного интеллекта.

Цифровые проекты в американских библиотеках

Безусловно интересным представляется опыт американских библиотек, экспериментирующих с подготовкой коллекций как структур данных, разрабатывающих варианты использования этих коллекций и формулирующих рекомендации к формам функциональных требований для разработчиков программного обеспечения. Направление «цифровые коллекции как данные» активно развивается в университетских библиотеках, привлекают внимание проекты Библиотеки Конгресса США, Нью-Йоркской публичной библиотеки и библиотек штатов. В настоящей статье мы не будем останавливаться на этих проектах, а охарактеризуем важные национальные инициативы, послужившие институционализации направления «цифровые коллекции как данные» и способствовавшие формированию междисциплинарного профессионального сообщества. Именно библиотеки должны быть опорными базами проведения научных исследований на основе своих цифровых массивов.

В 2016 г. Американская библиотечная ассоциация подготовила специальный отчет «Цифровые гуманитарные науки в библиотеках» [27], в котором подчеркивалось, что они оказывают значительное влияние на академический ландшафт уже более десяти лет. Это не может не учитываться в работе современных библиотек, которые должны стать активными участниками инновационных проектов.

В 2016–2018 гг. в США при поддержке спонсорских организаций был реализован проект «Всегда уже вычислительный: коллекции как данные» [28]. Целью проекта являлось обобщение опыта и имеющихся практик библиотек для выявления основных проблем и тех областей деятельности, которые должны быть углубленно изучены. В течение двух лет специалисты анализировали имеющиеся регламентирующие документы, обсуждали существующие и потенциально перспективные подходы к подготовке коллекций для использования в научных исследованиях и процессах обучения при помощи специальных программно-технологических средств. В рамках проекта было проведено два национальных форума, организовано несколько семинаров, о результатах проекта рассказывалось на междисциплинарных и профессиональных конференциях, подготовлено около десятка материалов, которые должны стать руководством для учреждений, рассматривающих возможность создания коллекций как данных. Но главное, в рамках проекта была сделана попытка разработки национальной стратегии работы с коллекциями как данными, стратегических планов библиотек, в которых приоритетное внимание уделялось работе с коллекциями как данными и многое другое. Обсуждение результатов проекта не ограничивалось только площадками библиотечного сообщества, также оно проводилось в Обществе американских архивистов (Society of American Archivists), Федерации цифровых библиотек (Digital Library Association), Американской исторической ассоциации (American History Association) и др. Именно это позволило учесть все возможные критические замечания и пожелания специалистов смежных областей.

Разработка тематики была продолжена в 2019–2021 гг. в проекте «Коллекции как данные: от части к целому» [29]. В течение трех лет 12 проектных групп разрабатывали модели для превращения коллекций в данные, занимались реконцептуализацией услуг, поддерживающих научное использование библиотечных ресурсов. Участниками проекта стали преимущественно университеты (Университет штата Аризона, Гарвардский университет и др.) и Музей искусства Карнеги (Carnegie Museum of Art). После завершения проектов был сделан вывод, что результаты проектной деятельности обладают высокой исследовательской ценностью, представляя различные типы контента, языков и описательных практик. Помимо это-

го, участники проектов пришли к четкому пониманию сущности деятельности и определили ее базовые принципы, которые были изложены в двух документах-положениях: «Заявление Санта-Барбары о коллекциях как данных» [30] и «Ванкуверское заявление о коллекциях как данных» [31]. Для обсуждения был предложен набор базовых принципов, на которые должны опираться специалисты, работающие в рассматриваемом направлении. Появление таких документов показывает, что зарубежные коллеги, осуществляя переход от конкретных институциональных проектов, формируют методологические основы деятельности. Поскольку упомянутые документы открыты для обсуждения и пока не носят заверщенного характера, приведем краткую характеристику лишь отдельных принципов общего порядка, которые представляются важными.

- Разработки по направлению «цифровые коллекции как данные» направлены на стимулирование использования коллекций оцифрованных материалов и исходно цифровых документов при помощи информационных технологий. Превращение коллекции из статичных документов в наборы данных помогают расширить возможности взаимодействия читателей с коллекциями.

- Исключается историческое и социальное неравенство, которое может проявляться в получении доступа к коллекциям, их объеме, описании и использовании. Обязательства по борьбе с неравенством должны быть отражены в соответствующих документах и обнародованы.

- Технологии для работы с цифровыми коллекциями должны быть максимально доступны за исключением тех случаев, когда это не представляется возможным по причинам этического и юридического характера.

- Подходы к разработке данных и доступу к ним могут быть различными, включая подготовку общедоступной документации: стратегических планов, должностных инструкций, схем метаданных, считываемых человеком и машиной, таблиц данных, технологических процессов и др.

Итоги анализа зарубежного опыта и задачи отечественных библиотек

Подводя итоги анализа зарубежного опыта, необходимо отметить, что направление деятельности библиотек «цифровые коллекции

как данные» относится к приоритетам развития библиотечных учреждений в целом. Это подтверждается включением тематики в основополагающие документы, касающиеся в том числе стратегии комплектования, организации доступа, создания новых видов обслуживания. С разной степенью интенсивности направление развивается во всех типах библиотек — публичных, университетских, национальных. На базе библиотек создаются центры цифровых гуманитарных наук, позволяющие привлекать к новым разработкам специалистов различного профиля. Развитие направления «цифровые коллекции как данные» идет по двум направлениям: для организации углубленного доступа и научных исследований; для обучения пользователей. Публикация данных осуществляется как на институциональных, так и на сторонних платформах. Это позволяет значительно расширить аудиторию. Предоставление данных не ограничивается только организационными и технологическими проблемами, но также подразумевает учет требований законодательства, в том числе авторского права. По мнению специалистов, реализация проектов «коллекции как данные» требует скорее организационных, а не технических усилий. В настоящее время идет процесс становления направления, анализ которого еще предстоит сделать в дальнейшем, в частности в контексте технологических и форматных аспектов.

В России использование информационных технологий для повышения эффективности гуманитарных исследований, преимущественно исторических, рассматривается в контексте исторической информатики. Большое значение для консолидации сообщества принесла Ассоциация «История и компьютер», созданная в 1992 г., а также проводимые при поддержке этой ассоциации конференция и научные исследования. Исследователи отмечали, что «с появлением больших массивов оцифрованных исторических текстов, обогащенных метаданными, связанными с историческим контекстом, стало возможным проводить более качественный анализ текстов. Анализ изобразительных источников также может быть значительно углублен за счет применения исследовательских методов, основанных на контексте, т. е. с использованием контекстной информации, включенной в метаданные» [32].

С 2012 г. издается электронный журнал «Историческая информатика», а в последние годы появляются сетевые ресурсы непосред-

ственно по цифровым гуманитарным наукам. В 2016 г. вышел в свет сборник статей «Роль библиотек в информационном обеспечении исторической науки» [33], в публикациях которого не только обоснована необходимость, но и доказана возможность использования ресурсов библиотек при компьютерных исследованиях. В частности, был описан опыт сотрудничества Пермского государственного национального исследовательского университета, Пермской государственной краевой универсальной библиотеки им. А.М. Горького и библиотеки Пермского краеведческого музея по реализации проекта «Российские парламентарии начала XX века»⁷, предусматривавшего создание одноименной информационной системы на основе данных указателей к стенографическим отчетам Государственной думы Российской империи. В рамках сотрудничества с Российской государственной библиотекой (РГБ) участники проекта для создания информационных систем по парламентской истории России получили ценные биографические сведения о депутатах Государственной думы и членах Государственного совета. Это позволило уточнить место жительства, род занятий членов российского парламента после революции, даты и место их смерти. Серия проектов была реализована на основе фондов библиотеки Пермского краеведческого музея с применением новых компьютерных технологий [34].

Из смежных областей также можно привести ряд примеров. Отметим результат деятельности Лаборатории цифровых исследований литературы и фольклора Института русской литературы РАН — репозиторий открытых данных по русской литературе и фольклору, созданный в том числе и для расширения количественных исследований в литературоведении. Самым близким нам по времени стал проект «Digital Петр», в рамках которого в 2022 г. был открыт для доступа интернет-ресурс «Digital Петр» с программой по машинному чтению неопубликованных и опубликованных автографов Петра Великого с комментариями к ним [35]. На базе Сибирского федерального университета представлена демонстрационная версия платформы «Сибириана»⁸ для исследований исторического, культурного и природного наследия Енисейской Сибири.

В 2023 г. вышла в свет первая отечественная коллективная монография, посвященная цифровым гуманитарным исследованиям и новым способам работы с оцифрованным культурным наследием [36]. Ведущие отече-

ственные университеты готовят кадры в области цифровых гуманитарных наук (например, Пермский государственный национальный исследовательский университет, Национальный исследовательский университет ИТМО, Сибирский федеральный университет, Алтайский государственный гуманитарно-педагогический университет им. В.М. Шукшина, Томский государственный университет, Национальный исследовательский университет «Высшая школа экономики», Балтийский федеральный университет им. И. Канта). В рамках университетских образовательных программ реализуются научные проекты по моделированию исторических процессов, компьютеризованному анализу текстов исторических источников, исторической геоинформатике и др.

В Интернете можно найти различного рода практические пособия, в частности на образовательном портале «Системный Блок»⁹: «Как распознать тексты и сделать корпус для исследования: пошаговая инструкция», «Как сделать тематическое моделирование», «Корпус из твитов своими руками» и др.

В качестве положительного момента можно отметить экспериментальные разработки РГБ в области использования нейросетей и искусственного интеллекта в работе с книгами и информацией [37; 38], применение программ по выявлению совпадений в текстах с источниками полнотекстовых коллекций Электронной библиотеки РГБ («Антиплагиат»). Насколько известно, научные разработки в области компьютерной датафикации ведутся в Российской национальной библиотеке. К проблеме необходимости использования искусственного интеллекта обращаются ведущие российские библиотекovedы [39; 40].

Таким образом, в последние десятилетия в стране сформировалась научно-технологическая среда для поддержки исследований на базе ресурсов библиотек, которые должны не только преодолеть свою оторванность от реальных научных исследований в сфере цифровых гуманитарных наук, но и войти в нее в качестве равноправного партнера. С учетом того, что практически все цифровые коллекции, создаваемые в российских библиотеках, предназначались для применения в образовании и как поддержка науки, потенциал их дальнейшего использования очень велик. Обозначим задачи, которые встанут перед библиотеками.

1. Создание в структуре библиотек специализированных центров по изучению собствен-

ных цифровых ресурсов на предмет целесообразности их представления в виде данных в контексте потенциальной аудитории, качества имеющегося цифрового контента, наличия расширенного набора метаданных, отражающих как отдельные документы, так и коллекции в целом. Тенденция создания лабораторий на базе библиотек нашла отражение в отечественной практике. Например, в 2014 г. на базе научной библиотеки Томского государственного университета была создана Лаборатория библиотечных и коммуникативных исследований¹⁰, в том числе для создания проблемно-тематических ресурсов с современным аналитическим инструментарием. Один из значимых проектов — это создание конструктора комбинаторных поисковых запросов по истории и социальным наукам [41]. В функции вновь создаваемых центров должны входить также подготовка данных и определение структур наборов данных, форматов, решение вопросов лицензирования и выявления ограничений на повторное использование данных коллекций.

2. Повышение качества оцифровки документов из библиотечных фондов, в частности в целях их массового распознавания.

3. Привлечение к деятельности «коллекции как данные» в библиотеках представителей различных профессиональных сообществ (исторического, филологического, компьютерных наук) из академической и, прежде всего, университетской науки. Библиотеки должны стать «координационной структурой для гуманитарных исследований и проектных практик, в том числе онлайн-взаимодействия ученых» [42, с. 520]. Это касается не только подключения библиотек к уже ведущимся в других сферах проектам, но и инициирование и активное развитие собственных проектных решений.

4. Обучение не только персонала, но и пользовательской аудитории, ее подготовка к применению программно-технологических средств — инструментов, позволяющих работать с цифровыми коллекциями как данными.

Примечания

- ¹ Library Labs [Online-Event] // Zentralbibliothek Zürich. URL: <https://www.zb.uzh.ch/de/events/library-labs> (дата обращения: 26.01.2024).
- ² Living Knowledge — a new vision for the British Library 2015-23 // YouTube. British Library. URL: <https://www.youtube.com/watch?v=HB2hKfJxRrc> (дата обращения: 26.01.2024).

- ³ Rethinking Research to Illuminate Our Past // Living with Machines. URL: <https://livingwithmachines.ac.uk/> (дата обращения: 26.01.2024).
- ⁴ British Library Datasets // British Library's Research Repository. URL: <https://bl.iro.bl.uk/collections/64e3804a-788a-4c4b-962c-ae180d955455?locale=en> (дата обращения: 26.01.2024).
- ⁵ Search from Digital Materials // National Library of Finland. URL: https://digi.kansalliskirjasto.fi/etusivu?set_language=en (дата обращения: 26.01.2024).
- ⁶ Welcome to the Greek National Aggregator of Digital Cultural Content // SearchCulture.gr. URL: <https://www.searchculture.gr/aggregator/portal/?language=en> (дата обращения: 26.01.2024).
- ⁷ Российские парламентарии начала XX века // Центр цифровой гуманитаристики Пермского университета. URL: <http://dh.psu.ru/project/03-01-12012/> (дата обращения: 26.01.2024).
- ⁸ Исследуй историческое, культурное и природное наследие Енисейской Сибири / Сибириана: агрегатор наследия. URL: <https://siberiana.online/> (дата обращения: 26.01.2024).
- ⁹ Системный Блок : онлайн-журнал. URL: <https://sysblok.ru/> (дата обращения: 26.01.2024).
- ¹⁰ Лаборатория библиотечных и коммуникативных исследований. URL: <https://project.lib.tsu.ru/liblabtsu> (дата обращения: 26.01.2024).

Список источников

1. Кижнер И.А. Цифровые технологии культурного наследования в современном обществе : дис. ... канд. культурологии. Красноярск, 2020. 182 с. URL: <https://research.sfu-kras.ru/node/14053> (дата обращения: 26.01.2024).
2. Савицкая Т.Е. Новые библиотечные услуги в рамках проектов цифровых гуманитарных наук: зарубежный опыт // Библиотековедение. 2021. Т. 70, № 1. С. 55—64. DOI: 10.25281/0869-608X-2021-70-1-55-64.
3. Юмашева Ю.Ю. Историческая наука, библиотеки, архивы, музеи и искусственный интеллект: что день грядущий нам готовит? // Документ. Архив. История. Современность : сборник научных трудов / [гл. ред. Л.Н. Мазур]. Екатеринбург : Изд-во Урал. ун-та, 2021. № 21. С. 247—279. URL: <https://elar.urfu.ru/handle/10995/104227> (дата обращения: 26.01.2024).
4. Жабко Е.Д. Интеграция ресурсов библиотек, музеев и архивов: вместе или самостоятельно? // Биб-

- лиотековедение. 2023. Т. 72, № 2. С. 103–115. DOI: 10.25281/0869-608X-2023-72-2-103-115.
5. *Padilla T.G.* Collections as Data. Implications for Enclosure // *College and Research Libraries News*. 2018. Vol. 79, № 6. P. 296–300. URL: <https://crln.acrl.org/index.php/crlnews/article/view/17003/18751> (дата обращения: 10.01.2024).
6. *Wittmann R., Neatrou A., Cummings R., Myntti J.* From Digital Library to Open Datasets : Embracing a “Collections as Data” Framework // *Information Technology and Libraries*. 2019. № 38 (4). P. 49–61. DOI: 10.6017/ital.v38i4.11101.
7. *Padilla T., Kettler H.S., Shorish Y.* Collection as Data: Part to Whole Final Report // *Zenodo*. 2023. 20 November. URL: <https://zenodo.org/records/10161976> (дата обращения: 08.01.2024).
8. *Neely L., Luther A., Weinard C.* Cultural Collections as Data: Aiming for Digital Literacy and Tool Development // *MuseWeb Conference 2019* (Boston, 2–6 April 2019). URL: <https://mw19.mwconf.org/paper/cultural-collections-as-data-aiming-for-digital-data-literacy-and-tool-development/index.html> (дата обращения: 30.01.2024).
9. *Mordell D.* Critical Questions for Archives as (Big) Data // *The Journal of the Association of Canadian Archivists*. 2019. № 87. P. 140–161. URL: <https://www.archivaria.ca/index.php/archivaria/article/view/13673/15071> (дата обращения: 30.01.2024).
10. *Neudecker C.* Building Library Labs — What Do They Do and Who Are They for? October 10, 2018 // *EuropeanaPro*. URL: <https://pro.europeana.eu/post/building-library-labs-what-do-they-do-and-who-are-they-for> (дата обращения: 30.01.2024).
11. Бёрлесон Л., Хаффенден К., Мальмстен М., Клигволл Ф., Ренде Э., Курц Р., Рекатати Ф., Хэгллэф Х., Сикора Ю. Преобразование библиотек в цифровую исследовательскую инфраструктуру: создание KBLab в Национальной библиотеке Швеции // *Новости Международной федерации библиотечных ассоциаций и учреждений*. 2023. № 2. С. 37–52. URL: <https://novostifla.rsl.ru/article/view/788> (дата обращения: 30.01.2024).
12. *Савицкая Т.Е.* Сдвиг парадигм: библиотеки в контексте цифровых гуманитарных наук (зарубежный опыт) // *Обсерватория культуры*. 2018. Т. 15, № 5. С. 532–541. DOI: 10.25281/2072-3156-2018-15-5-532-541.
13. *Mahey M., Al-Abdulla A., Ames S., Bray P., Candela G., Chambers S., Derven C., Dobrev-McPherson M., Gasser K., Karner S., Kokegei K., Laursen D., Potter A., Straube A., Wagner S.-C., Wilms L.* Open a GLAM Lab. Doha, Qatar : Book Sprint, 2019. 166 p. URL: <https://glamlabs.pubpub.org/> (дата обращения: 30.01.2024).
14. *Ames S.* Transparency, Provenance and Collections as Data: The National Library of Scotland’s Data Foundry // *LIBER Quarterly : The Journal of the Association of European Research Libraries*. 2021. Vol. 31, № 1. P. 1–13. DOI: 10.18352/lq.10371.
15. *Ames S.* Special Collections as Data: The National Library of Scotland’s Data Foundry // *Consortium of European Research Libraries*. URL: <https://cerlblog.wordpress.com/2020/08/13/special-collections-as-data-the-national-library-of-scotlands-data-foundry/> (дата обращения: 30.01.2024).
16. *Farquhar A.* Digital scholarship and the Cloud / British Library : [презентация] // *The 9th International Digital Curation Conference (IDCC)*. San Francisco, 27 February 2014. 33 p. URL: <https://www.dcc.ac.uk/events/idcc14/workshops> (дата обращения: 30.01.2024).
17. *Jones J.* The British Library Puts Over 1,000,000 Images in the Public Domain: A Deeper Dive into the Collection // *Open Culture*. 2015. 8 September. URL: <http://www.openculture.com/2015/09/the-british-library-puts-over-1000000-images-in-the-public-domain-a-deeper-dive-into-the-collection.html> (дата обращения: 30.01.2024).
18. *Kotarski R., Reimer T.* Making Everything Available: British Library Research Services and Research Data Strategy // *International Journal of Digital Curation*. 2018. Vol. 13, № 1. P. 161–169. DOI: 10.2218/ijdc.v13i1.605.
19. *British Library Annual Report and Accounts 2022/23*. London : British Library, 2022. 104 p. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1172034/British_Library_Annual_Report_and_Accounts_2022-23_FINAL.pdf (дата обращения: 30.01.2024).
20. *Liem J., Goudarouli E., Hirschon S., Wood J., Perin C.* Geoblobs and World War One // *CharlesPerin.net*. URL: http://charlesperin.net/projects/war_diaries (дата обращения: 30.01.2024).
21. *Beals M., Bell E.* The Atlas of Digitised Newspapers and Metadata : Reports from Oceanic Exchanges. Loughborough : Loughborough University Publ., 2020. 180 p. DOI: 10.6084/m9.figshare.11560059.
22. *Раутайнен Ю.* Библиотечные электронные коллекции финских газет в цифровых гуманитарных исследовательских проектах / пер. А.Е. Зуев // *Библиотековедение*. 2023. Т. 72, № 1. С. 51–61. DOI: 10.25281/0869-608X-2023-72-1-51-61.
23. *Late E., Kumpulainen S.* Interacting with Digitised Historical Newspapers: Understanding the Use of Digital Surrogates as Primary Sources // *Journal of Documentation*. 2022. Vol. 78, № 7. P. 106–124. DOI: 10.1108/JD-04-2021-0078.

24. *Chambers S., Lemmers F.* Experimenting with Collections as Data at KBR, Royal Library of Belgium: Exploring Sustainable Workflows to Facilitate Corpus Building in the Digital Humanities // Историко-культурное наследие в цифровом измерении : материалы Международной научной конференции (г. Пермь, 20–22 октября 2021 г.) / Пермский государственный национальный исследовательский университет. Пермь, 2021. С. 20–24.
25. *Strategy 2020–2025. Empowering Digital Chance.* Luxembourg : Publications Office of the European Union, 2020. 48 p. URL: https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Strategy%202020%20-%202025.pdf (дата обращения: 30.01.2024).
26. *Candela G., Gabriëls N., Chambers S., Pham T.-A., Ames S., Fitzgerald N., Hofmann K., Harbo V., Potter A., Ferriter M., Manchester E., Irollo A., Van Keer E., Mahey M., Holownia O., Dobрева M.* A Checklist to Publish Collections as Data in GLAM Institution // Cornell University. Computer Science. Digital Libraries. URL: <https://arxiv.org/abs/2304.02603v2> (дата обращения: 30.01.2024).
27. *Varner S., Hswe P.* Special Report: Digital Humanities in Libraries // American Libraries. 2016. 4 January. URL: <https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/> (дата обращения: 30.01.2024).
28. *Padilla T., Allen L., Frost H., Potvin S., Russey Roke E., Varner S.* Always Already Computational: Collections as Data : Final Report. 2019. 22 May. DOI: 10.17605/OSF.IO/MX6UK.
29. *Collections as Data : Part to Whole.* URL: <https://collectionsasdata.github.io/part2whole/> (дата обращения: 30.01.2024).
30. *The Santa Barbara Statement on Collections as Data (2017). Version 2* // Always Already Computational – Collections as Data. URL: <https://collectionsasdata.github.io/statement/> (дата обращения: 30.01.2024).
31. *Padilla T., Scates K., Varner S.* Vancouver Statement on Collections as Data // Zenodo. 2023. 13 September. DOI: 10.5281/zenodo.8341519.
32. *Гарскова И.М.* Историческая информатика: точки бифуркации // Круг идей: модели и технологии исторических реконструкций : труды XI конференции Ассоциации «История и компьютер» / под ред. Л.И. Бородинки, В.Н. Владимирова, Г.В. Можяевой. Москва ; Барнаул; Томск : Изд-во Московского университета, 2010. С. 5–33.
33. *Роль библиотек в информационном обеспечении исторической науки : сборник статей / авт.-сост. Е.А. Воронцова ; отв. ред. А.О. Чубарьян, В.Р. Фирсов.* Москва : Этерна, 2016. 672 с.
34. *Корниенко С.И., Поврозник Н.Г., Гагарина Д.А., Харисова А.Р.* Университет и библиотека: сотрудничество в области сохранения и изучения губернской периодики // Роль библиотек в информационном обеспечении исторической науки : сборник статей / авт.-сост. Е.А. Воронцова ; отв. ред. А.О. Чубарьян, В.Р. Фирсов. Москва : Этерна, 2016. С. 500–515.
35. *Базарова Т., Димитров Д., Потанин М., Прокуракова М.* Распознать и транскрибировать: автографы Петра Великого и технологии искусственного интеллекта // Воронцово поле. 2020. № 4. С. 64–71.
36. *Цифровые гуманитарные исследования : монография / А.Б. Антопольский, А.А. Бонч-Осмоловская, Л.И. Бородин [и др.] ; Сибирский федеральный университет.* Красноярск, 2023. 272 с.
37. *Дуда В.В.* Искусственный интеллект на службе у книг: традиции и инновации главной библиотеки страны : [интервью с генеральным директором Российской государственной библиотеки] / беседу вела В. Галкина // Литературная газета. 2022. № 22 (6836). URL: <https://lgz.ru/article/-22-6836-01-06-2022/iskusstvennyy-intellektna-sluzhbe-u-knig/> (дата обращения: 30.01.2024).
38. *Дуда В.В.* Библиотека в пространстве социальных трансформаций: сохранение культурного наследия и вызовы цифровизации / беседу вела: Е.В. Никонорова, Е.А. Шибяева // Библиотекосведение. 2024. Т. 73, № 1. С. 7–22. DOI: 10.25281/0869-608X-2024-73-1-7-22.
39. *Столяров Ю.Н.* Искусственный интеллект и книжная библиотечная отрасль: направления разработки проблемы // Научные и технические библиотеки. 2022. № 1. С. 17–34. DOI: 10.33186/1027-3689-2022-1-17-34.
40. *Нещерет М.Ю.* Нейросети в библиотеке: новое в библиографическом обслуживании // Научные и технические библиотеки. 2024. № 1. С. 105–128. DOI: 10.33186/1027-3689-2024-1-105-128.
41. *Бочаров А.В.* Визуальный конструктор пользовательских поисковых запросов на основе комбинаторных тематических тезаурусов: реализация идей // Историческая информатика : Информационные технологии и математические методы в исторических исследованиях и образовании. 2014. № 2–3. С. 10–20. URL: https://kleio.asu.ru/2014/2-3/hcsj-232014_contents.html (дата обращения: 30.01.2024).
42. *Астафьева О.Н., Никонорова Е.В., Шлыкова О.В.* Культура в цифровой цивилизации: новый этап осмысления стратегии будущего для устойчивого развития // Обсерватория культуры. 2018. Т. 15, № 5. С. 516–531. DOI: 10.25281/2072-3156-2018-15-5-516-531.

Digital Collections as Data: A New Direction for Libraries

Elena D. Zhabko,

Information Historical Research Centre — Military-Historical Library of the General Staff of the Armed Forces of the Russian Federation, 10 Palace Sq., St. Petersburg, 191055, Russia
ORCID 0000-0002-0871-5874; SPIN 9942-9780; edzhabko@gmail.com

Abstract. The aim of the article is to analyze the theoretical foundations and experience of foreign library institutions in the direction of “digital collections as data”. Digital collections are considered not only as new forms of knowledge organization, but also as a tool for conducting scientific research in the field of humanities disciplines. Practice has shown that huge arrays of digitized texts and images can be used more successfully if modern information technologies are applied. Combining research methodology in the humanities and social sciences and computer tools makes it possible to extract quantitative and geographical data, facts, images on the basis of automatic analysis of recognized texts, visualize and display data in various formats, provide detailed metadata, model historical processes and much more. The most illustrative examples of innovative projects in a number of European libraries are given. The current state of transformation of digital collections as data in the United States is briefly characterized. It is determined that in recent decades in Russia a scientific, educational and technological environment for research based on library resources has been formed. Libraries should overcome their detachment from scientific research in the field of digital humanities and enter it as an equal partner. The tasks of domestic libraries for further development of digital content are defined, taking into account the achievements in the field of information technologies and the current level of technical knowledge.

Key words: libraries, archives, museums, digital collections as data, digital humanities laboratories, digital collections, information technology, open data, foreign experience, library and information services, national libraries, libraries of educational institutions.

Citation: Zhabko E.D. Digital Collections as Data: A New Direction for Libraries, *Bibliotekovedenie* [Russian Journal of Library Science], 2024, vol. 73, no. 2, pp. 129–142. DOI: 10.25281/0869-608X-2024-73-2-129-142.

References

1. Kizhner I.A. Digital Technologies of Cultural Inheritance in Modern Society, *Sibirskii federal'nyi universitet* [Siberian Federal University], Cand. cult. sci. diss. Krasnoyarsk, 2020, 182 p. Available at: <https://research.sfu-kras.ru/node/14053> (accessed 26.01.2024) (in Russ.).
2. Savitskaya T.E. New Library Services in the Framework of Digital Humanities Projects: Foreign Experience, *Bibliotekovedenie* [Russian Journal of Library Science], 2021, vol. 70, no. 1, pp. 55–64. DOI: 10.25281/0869-608X-2021-70-1-55-64 (in Russ.).
3. Yumasheva Yu.Yu. Historical Science, Archives, Libraries, Museums and Artificial Intelligence: What Does Tomorrow Hold? *Dokument. Arkhiv. Istoriya. Sovremennost': sbornik nauchnykh trudov* [Document. Archive. History. Modernity: collected articles]. Yekaterinburg, Izd-vo Ural. Un-ta, 2021, no. 21, pp. 247–279. Available at: <https://elar.urfu.ru/handle/10995/104227> (accessed 26.01.2024) (in Russ.).
4. Zhabko E.D. Integration of Library, Museum and Archive Resources: Together or Independently? *Bibliotekovedenie* [Russian Journal of Library Science], 2023, vol. 72, no. 2, pp. 103–115. DOI: 10.25281/0869-608X-2023-72-2-103-115 (in Russ.).
5. Padilla T.G. Collections as Data. Implications for Enclosure, *College and Research Libraries News*, 2018, vol. 79, no. 6, pp. 296–300. Available at: <https://crln.acrl.org/index.php/crlnews/article/view/17003/18751> (accessed 10.01.2024).
6. Wittmann R., Neatrou A., Cummings R., Myntti J. From Digital Library to Open Datasets: Embracing a “Collections as Data” Framework, *Information Technology and Libraries*, 2019, no. 38 (4), pp. 49–61. DOI: 10.6017/ital.v38i4.11101.
7. Padilla T., Kettler H.S., Shorish Y. Collection as Data: Part to Whole Final Report, *Zenodo*. 2023, November 20. Available at: <https://zenodo.org/records/10161976> (accessed 08.01.2024).

8. Neely L., Luther A., Weinard C. Cultural Collections as Data: Aiming for Digital Literacy and Tool Development, *MuseWeb Conference 2019 (Boston, April 2–6, 2019)*. Available at: <https://mw19.mwconf.org/paper/cultural-collections-as-data-aiming-for-digital-data-literacy-and-tool-development/index.html> (accessed 30.01.2024).
9. Mordell D. Critical Questions for Archives as (Big) Data, *The Journal of the Association of Canadian Archivists*, 2019, no. 87, pp. 140–161. Available at: <https://www.archivaria.ca/index.php/archivaria/article/view/13673/15071> (accessed 30.01.2024).
10. Neudecker S. Building Library Labs – What Do They Do and Who Are They for? *EuropeanaPro*. 2018, October 10. Available at: <https://pro.europeana.eu/post/building-library-labs-what-do-they-do-and-who-are-they-for> (accessed 30.01.2024).
11. Börjeson L., Haffenden Ch., Malmsten M., Klingwall F., Rende E., Kurts R., Rekathati F., Hägglöf H., Sikora J. Transfiguring the Library as Digital Research Infrastructure Making Kblab at the National Library of Sweden, *Novosti Mezhdunarodnoi federatsii biblioteknykh assotsiatsii i uchrezhdenii* [Newsletter of the International Federation of Library Associations and Institutions], 2023, no. 2, pp. 37–52. Available at: <https://novostifla.rsl.ru/article/view/788> (accessed 30.01.2024) (in Russ.).
12. Savitskaya T.E. Paradigm Shift: Libraries in the Context of Digital Humanities (Foreign Experience), *Observatoriya kul'tury* [Observatory of Culture], 2018, vol. 15, no. 5, pp. 532–541. DOI: 10.25281/2072-3156-2018-15-5-532-541 (in Russ.).
13. Mahey M., Al-Abdulla A., Ames S., Bray P., Candela G., Chambers S., Derven C., Dobрева-McPherson M., Gasser K., Karner S., Kokegei K., Laursen D., Potter A., Straube A., Wagner S.-C., Wilms L. *Open a GLAM Lab*. Doha, Qatar, Book Sprint Publ., 2019, 166 p. Available at: <https://glamlabs.pubpub.org/> (accessed 30.01.2024).
14. Ames S. Transparency, Provenance and Collections as Data: The National Library of Scotland's Data Foundry, *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 2021, vol. 31, no. 1, pp. 1–13. DOI: 10.18352/lq.10371.
15. Ames S. Special Collections as Data: The National Library of Scotland's Data Foundry, *Consortium of European Research Libraries*. Available at: <https://cerlblog.wordpress.com/2020/08/13/special-collections-as-data-the-national-library-of-scotlands-data-foundry/> (accessed 30.01.2024).
16. Farquhar A. Digital Scholarship and the Cloud, *The 9th International Digital Curation Conference (IDCC)*. San Francisco, 2014, February 27, 33 p. Available at: <https://www.dcc.ac.uk/events/idcc14/workshops> (accessed 30.01.2024).
17. Jones J. The British Library Puts Over 1,000,000 Images in the Public Domain: A Deeper Dive into the Collection, *Open Culture*. 2015, September 8. Available at: <http://www.openculture.com/2015/09/the-british-library-puts-over-1000000-images-in-the-public-domain-a-deeper-dive-into-the-collection.html> (accessed 30.01.2024).
18. Kotarski R., Reimer T. Making Everything Available: British Library Research Services and Research Data Strategy, *International Journal of Digital Curation*, 2018, vol. 13, no. 1, pp. 161–169. DOI: 10.2218/ijdc.v13i1.605.
19. *British Library Annual Report and Accounts 2022/23*. London, British Library Publ., 2022, 104 p. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1172034/British_Library_Annual_Report_and_Accounts_2022-23_FINAL.pdf (accessed 30.01.2024).
20. Liem J., Goudarouli E., Hirschon S., Wood J., Perin C. Geoblobs and World War One, *CharlesPerin.net*. Available at: <http://charlesperin.net/projects/war-diaries> (accessed 30.01.2024).
21. Beals M., Bell E. *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges*. Loughborough, Loughborough University Publ., 2020, 180 p. DOI: 10.6084/m9.figshare.11560059.
22. Rautiainen J., Zuev A.E. (transl.) Digitized Finnish Newspapers in Digital Humanities Research Projects: Challenges and Solutions from the Library Perspective, *Bibliotekovedenie* [Russian Journal of Library Science], 2023, vol. 72, no. 1, pp. 51–61. DOI: 10.25281/0869-608X-2023-72-1-51-61 (in Russ.).
23. Late E., Kumpulainen S. Interacting with Digitised Historical Newspapers: Understanding the Use of Digital Surrogates as Primary Sources, *Journal of Documentation*, 2022, vol. 78, no. 7, pp. 106–124. DOI: 10.1108/JD-04-2021-0078.
24. Chambers S., Lemmers F. Experimenting with Collections as Data at KBR, Royal Library of Belgium: Exploring Sustainable Workflows to Facilitate Corpus Building in the Digital Humanities, *Istoriko-kul'turnoe nasledie v tsifrovom izmerenii: materialy Mezhdunarodnoi nauchnoi konferentsii (g. Perm', 20–22 oktyabrya 2021 g.)* [Cultural Heritage in the Digital Dimension: Proceedings of the International Conference (Perm, October 20–22, 2021)]. Perm, 2021, pp. 20–24.
25. *Strategy 2020–2025. Empowering Digital Chance*. Luxembourg, Publications Office of the European Union, 2020, 48 p. Available at: https://pro.europeana.eu/files/Europeana_Professional/Publications/

- Europeana%20Strategy%202020%20-%202025.pdf (accessed 30.01.2024).
26. Candela G., Gabriëls N., Chambers S., Pham T.-A., Ames S., Fitzgerald N., Hofmann K., Harbo V., Potter A., Ferriter M., Manchester E., Irollo A., Van Keer E., Mahey M., Holownia O., Dobрева M. A Checklist to Publish Collections as Data in GLAM Institution, *Cornell University. Computer Science. Digital Libraries*. Available at: <https://arxiv.org/abs/2304.02603v2> (accessed 30.01.2024).
 27. Varner S., Hswe P. Special Report: Digital Humanities in Libraries, *American Libraries*. 2016, January 4. Available at: <https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/> (accessed 30.01.2024).
 28. Padilla T., Allen L., Frost H., Potvin S., Russey Roke E., Varner S. *Always Already Computational: Collections as Data: Final Report*. 2019, May 22. DOI: 10.17605/OSF.IO/MX6UK.
 29. *Collections as Data: Part to Whole*. Available at: <https://collectionsasdata.github.io/part2whole/> (accessed 30.01.2024).
 30. The Santa Barbara Statement on Collections as Data. Version 2, *Always Already Computational – Collections as Data*. Available at: <https://collectionsasdata.github.io/statement/> (accessed 30.01.2024).
 31. Padilla T., Scates K., Varner S. Vancouver Statement on Collections as Data. *Zenodo*. 2023, September 13. DOI: 10.5281/zenodo.8341519.
 32. Garskova I.M. Historical Informatics: Bifurcation Points, *Krug idei: modeli i tekhnologii istoricheskikh rekonstruktsii: trudy XI konferentsii Assotsiatsii "Istoriya i komp'yuter"* [Circle of Ideas: Models and Technologies of Historical Reconstructions: Proceedings of the 11th Conference of the Association "History and Computer"]. Moscow, Barnaul, Tomsk, Izd-vo Moskovskogo Universiteta, 2010, pp. 5–33 (in Russ.).
 33. Vorontsova E.A. (ed.) *Rol' bibliotek v informatsionnom obespechenii istoricheskoi nauki: sbornik statei* [Role of Libraries in Information Support of Historical Science: collected articles]. Moscow, Ehterna Publ., 2016, 672 p.
 34. Kornienko S.I., Povroznik N.G., Gagarina D.A., Kharisova A.R. University and Library: Cooperation in the Preservation and Research of Provincial Periodicals, *Rol' bibliotek v informatsionnom obespechenii istoricheskoi nauki: sbornik statei* [Role of Libraries in Information Support of Historical Science: collected articles]. Moscow, Ehterna Publ., 2016, pp. 500–515 (in Russ.).
 35. Bazarova T., Dimitrov D., Potanin M., To Recognize and Transcribe: The Autographs of Peter the Great and Artificial Intelligence Technologies, *Vorontsovo pole* ["Vorontsovo Pole" Journal], 2020, no. 4, pp. 64–71 (in Russ.).
 36. Antopolsky A.B., Bonch-Osmolovskaya A.A., Borodkin L.I. et al. (eds.) *Tsifrovyye gumanitarnyye issledovaniya: monografiya* [Digital Humanities Research: monograph]. Krasnoyarsk, Sibirskii Federal'nyi Universitet Publ., 2023, 272 p.
 37. Duda V.V. Artificial Intelligence in the Service of Books: Traditions and Innovations of the Main Library of Russia, *Literaturnaya gazeta* [Literary Gazette]. 2022, no. 22 (6836). Available at: <https://lgz.ru/article/-22-6836-01-06-2022/iskusstvennyy-intellekt-na-sluzhbe-u-knig/> (accessed 30.01.2024) (in Russ.).
 38. Duda V.V., Nikonorova E.V., Shibaeva E.A. Library in the Space of Social Transformations: Preservation of Cultural Heritage and Challenges of Digitalization, *Bibliotekovedenie* [Russian Journal of Library Science], 2024, vol. 73, no. 1, pp. 7–22. DOI: 10.25281/0869-608X-2024-73-1-7-22 (in Russ.).
 39. Stolyarov Yu.N. Artificial Intellect and the Book Library Industry: Vectors for Problem Development, *Nauchnye i tekhnicheskie biblioteki* [Scientific and Technical Libraries], 2022, no. 1, pp. 17–34. DOI: 10.33186/1027-3689-2022-1-17-34 (in Russ.).
 40. Neshcheret M.Yu. Neural Networks in Libraries: A New Development in Bibliographic Services, *Nauchnye i tekhnicheskie biblioteki* [Scientific and Technical Libraries], 2024, no. 1, pp. 105–128. DOI: 10.33186/1027-3689-2024-1-105-128 (in Russ.).
 41. Bocharov A.V. The Visual Constructor of User's Web Search Queries Based on Combinatorial Thematic Thesauruses: Realization of the Idea, *Istoricheskaya informatika: Informatsionnye tekhnologii i matematicheskie metody v istoricheskikh issledovaniyakh i obrazovanii* [Historical Information Science: Information Technology and Quantitative Methods in Historical Research and Education], 2014, no. 2–3, pp. 10–20. Available at: https://kleio.asu.ru/2014/2-3/hcsj-232014_contents.html (accessed 30.01.2024) (in Russ.).
 42. Astafyeva O.N., Nikonorova E.V., Shlykova O.V. Culture in the Digital Civilization: A New Stage in Understanding the Future Strategy for Sustainable Development, *Observatoriya kul'tury* [Observatory of Culture], 2018, vol. 15, no. 5, pp. 516–531. DOI: 10.25281/2072-3156-2018-15-5-516-531 (in Russ.).