

УДК [025.173:004](480)
ББК 78.375.31(4Фин)с51
DOI 10.25281/0869-608X-2023-72-1-51-61

Ю. РАУТИАЙНЕН

Библиотечные электронные коллекции финских газет в цифровых гуманитарных исследовательских проектах*

Реферат. Библиотеки оказывают непосредственное влияние на исследовательские проекты в области цифровых гуманитарных наук (Digital Humanities, DH), в которых используются оцифрованные газеты из их фондов. Однако степень участия библиотеки в проекте может варьироваться — от поставщика ресурсов до активного партнера. С учетом различных требований для каждого проекта возникает необходимость в разнообразных библиотечных услугах.

Национальная библиотека Финляндии постоянно стремится повышать качество своих услуг и предоставлять ученым широкий доступ к оцифрованным коллекциям. Исследовательская группа DH библиотеки имеет большой опыт участия в научных проектах и активно поддерживает работу в области DH. В статье представлен обзор опыта исследовательской группы DH. Основное внимание уделено проектам, в которых использовались оцифрованные газеты и журналы. Описываются потенциальные проблемы и предлагаются возможные решения.

Ключевые слова: исследования, газеты, оцифровка, цифровые гуманитарные науки, библиотечно-информационное обслуживание.

Для цитирования: Раутиайнен 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.

Библиотека может играть различные роли в проектах, связанных с цифровыми гуманитарными науками (Digital Humanities, DH), например быть внешним источником данных, партнером по проекту и пр. Но независимо от своей роли, библиотека всегда будет оказывать влияние на проект. Исследовательская группа DH Национальной библиотеки Финляндии (НБФ) имеет много-

* Перевод выполнен в рамках деятельности Русскоязычного центра ИФЛА. Оригинал статьи см.: Rautiainen J. Digitized Finnish Newspapers in Digital Humanities Research Projects: Challenges and Solutions from the Library Perspective. URL: <https://repository.ifla.org/handle/123456789/1979>

Юха Раутиайнен,
Национальная библиотека Финляндии,
Научная библиотека,
специалист по информационным системам
Миккели, Финляндия
Saimaankatu, 6,
Mikkeli, 50100, Finland
E-mail: juha.rautiainen@helsinki.fi

летний опыт работы в проектах, использующих оцифрованные материалы. В ее состав входят руководитель отдела планирования, три специалиста по информационным системам и сотрудники, отвечающие за срочные проекты. Среди их ключевых обязанностей — поддержка ученых, участие в научных проектах, содействие доступу к оцифрованным материалам. Кроме того, исследовательская группа ДН тесно сотрудничает с несколькими другими подразделениями НБФ, в том числе со Службой оцифровки и исследований.

У цифровых гуманитарных наук есть много определений¹, и ученые из различных отраслей понимают этот термин по-разному. В целом можно сказать, что ДН работают на стыке гуманитарных наук и информатики.

В качестве вспомогательного средства в настоящем обзоре рассмотрены модели С. Обербихлер и соавторов [1], поскольку в них отражены те ожидания, которые возлагаются на библиотеки, участвующие в совместных исследованиях. В статье представлено несколько примеров различных проектов в области ДН, также эти модели выступают в качестве справочного материала для иллюстрации ряда проблем, которые могут возникнуть в ходе научного проекта. НБФ постоянно развивает свои услуги для удовлетворения потребностей пользователей. Автор рассказывает о нескольких ключевых решениях НБФ, некоторые были внедрены недавно, поэтому для их полной оценки требуется больше времени; однако в настоящей работе они приведены в качестве полезных примеров для тех, кто сталкивается с подобными вопросами.

Коллекция оцифрованных газет и система Digi

Оцифрованные газеты НБФ и новые газеты, получаемые в электронном виде, размещаются в Digi² (Цифровые коллекции НБФ). Digi содержит более 23 млн страниц оцифрованных материалов, в том числе более 14 млн газетных страниц. Каждый год добавляется около 2 млн страниц, и газеты составляют около половины нового оцифрованного контента.

В Digi работает свободный поиск по всем материалам, а к полному тексту всех ресурсов, не защищенных авторским правом, и некоторых ресурсов, защищенных авторским правом, можно получить доступ онлайн (либо свободно, либо войдя в систему как пользователь-иссле-

дователь). Все защищенные авторским правом материалы доступны только на рабочих станциях посетителей в библиотеках, получающих обязательный экземпляр. Некоторые из старейших оцифрованных ресурсов, не защищенных авторским правом, например оцифрованные финские газеты с 1771 по 1911 г., также доступны в виде загружаемых пакетов открытых данных и с помощью программного интерфейса приложения.

Согласно опросам, проведенным в 2018, 2019 и 2022 гг., Digi обслуживает различные пользовательские группы. Хотя в разных анкетах варианты ответа на вопрос о предполагаемом применении варьировались, специалисты по генеалогии всегда составляли самую многочисленную группу. Частные исследования или поиск и просмотр информации также являлись одними из самых популярных вариантов ответа, тогда как научная работа была на четвертом или шестом месте.

Модели междисциплинарного сотрудничества

Взаимопонимание между сторонами, участвующими в любом проекте, возникает в процессе общения, и разработка рабочего процесса для проекта выступает одним из инструментов, который может эффективно инициировать обмен идеями. Описание подробного рабочего процесса гарантирует, что необходимые задачи и компоненты действий станут понятными и «прозрачными», что поможет определить границы каждого шага в проекте и облегчит принятие сторонами компромиссных решений.

Я использую две модели, представленные в статье С. Обербихлер с коллегами [1], когда анализирую опыт исследовательской группы ДН; однако следует отметить, что данные модели не применялись напрямую ни в одном из указанных в этой статье проектов.

С. Обербихлер впервые описала конвейерную модель, ориентированную на оптимизацию рабочего процесса, для взаимодействия различных департаментов внутри библиотеки [1]. На рис. 1 показано долевое участие различных библиотечных отделов и применение (автоматизированных) вычислительных методов. Несмотря на то что у участников процесса общие задачи для работы с цифровыми коллекциями газет, они имеют разные цели и задействуют разные вычислительные методы.



Рис. 1. Представление междисциплинарного сотрудничества, ориентированного на оптимизацию рабочего процесса [1] (CC BY 4.0)

Например, добавление метаданных для предоставления контекста к материалу является в первую очередь технической задачей для специалиста по информатике, но это также интеллектуальная и трудоемкая задача для библиотекарей. Исследователи ДН оперируют данными, метаданными и выполняют качественный анализ, делая выводы, объединяющие все этапы работы.

Авторы [1] также предложили комплексный рабочий процесс цифровой интерпретации, который сочетает в себе дисциплинарные исследовательские подходы из компьютерных наук, гуманитарных наук и библиотечной деятельности. Эта в целом простая модель обобщает опыт и идеи, полученные в ходе междисциплинарного проекта NewsEye. Рабочий процесс для этой модели представлен на рис. 2.

Комплексная модель цифровой интерпретации рабочего процесса отличается от конвейерной модели тремя основными аспектами: подчеркивается роль данных, приоритет отдается итеративным (качественным) аналитическим шагам для постепенного получения более глубокого понимания данных, эта мо-

дель включает выявление возможных ошибок в обеих моделях, данные и инструменты, подчиненные реализации процессов, построенных на принципах герменевтики.

С точки зрения библиотек — хранителей культурного наследия — важно, чтобы каждое библиотечное подразделение знало, почему и как они отбирают материалы для оцифровки. Обмен метаданными имеет решающее значение, и каждый шаг процесса оцифровки должен отслеживаться.

Недавние и текущие проекты

Проекты, в которых используются оцифрованные газеты НБФ, разделены на три категории:

- исследователи работают независимо и получают доступ только к базовым услугам библиотеки;
- роль библиотеки заключается в первую очередь в наблюдении и консультировании;
- активное непосредственное участие библиотеки.

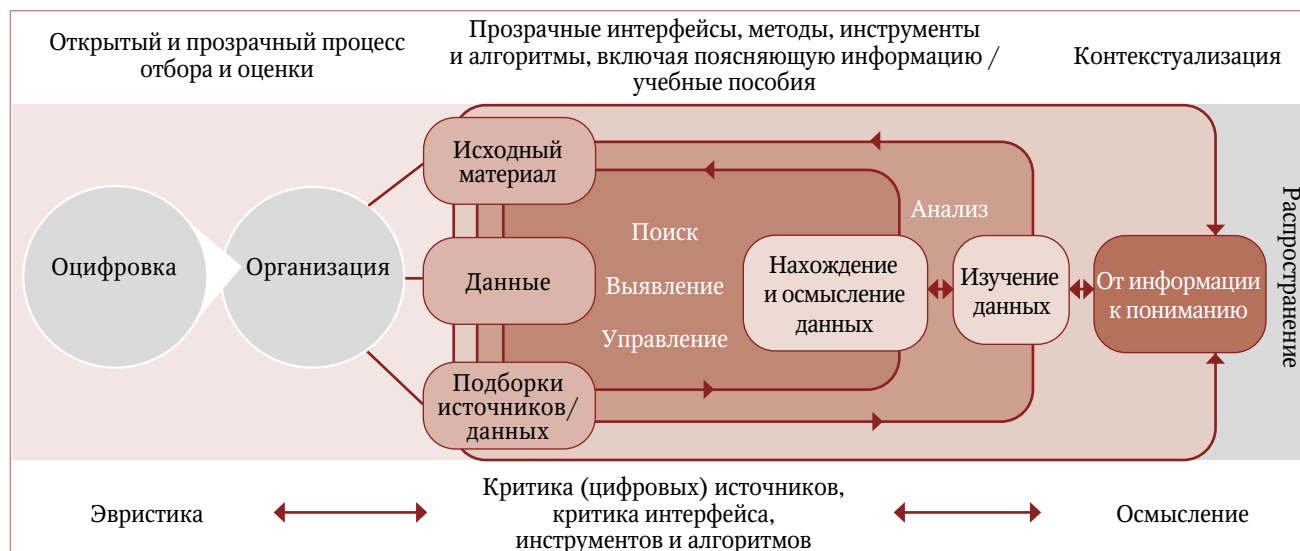


Рис. 2. Комплексная модель рабочего процесса цифровой интерпретации [1] (CC BY 4.0)

Независимые проекты. Можно с уверенностью предположить, что в большинстве проектов исследователи пользуются только стандартными услугами библиотеки. Ключевой проблемой с точки зрения библиотеки является то, что достоверную информацию об этих проектах и потребностях исследователей трудно получить. Поэтому совершенствование и развитие соответствующих услуг в некоторой степени основаны на допущениях.

Существует несколько способов получения информации о независимых проектах, однако это не всегда просто. Можно обращаться к опросам, чтобы напрямую узнавать о работе ученых, можно получать информацию из научных публикаций и новостей. Данные, собранные во время использования сервисов, также служат хорошим источником информации, кроме того, сами исследователи могут рассказывать библиотеке о своих проектах. Сведения, собранные с помощью этих процедур, зачастую носят отрывочный характер, но все равно они могут быть полезны для развития библиотечных услуг.

Первоначально наше внимание привлекла статья «Исторические тенденции весеннего таяния льда на реке Аура на юго-западе Финляндии за 1749–2018 гг.» [2]. В ней был представлен обновленный и дополненный набор данных о вскрытии льда в период 1749–2018 гг. на р. Аура в целях изучения климата. При нехватке статистических данных специалисты обращались к местным газетам как к дополнительному источнику информации. В данном случае все газеты имелись в открытом доступе онлайн, поэтому какая-либо прямая помощь со стороны НБФ не потребовалась.

Когда человек регистрируется в системе Digi в качестве пользователя-исследователя, ему необходимо заполнить короткую анкету, прежде чем он сможет получить доступ к материалам с ограниченным доступом. В анкете конкретно не запрашиваются сведения о проектах, однако многие респонденты добровольно предоставляют информацию о своей работе. Эти ответы играют важную роль для оценки перспектив применения оцифрованных газет в качестве исследовательского материала в Финляндии.

На сегодняшний день Digi имеет более 4,5 тыс. заполненных анкет, и оцифрованный материал часто выступает как традиционный библиотечный инструмент для просмотра газет или свободного поиска по тексту.

Однако были замечены и другие подходы. Например, в нескольких анкетах упоминался интеллектуальный анализ данных, а в одном проекте в области социальных наук посредством вычислительных методов сравнивались сведения, полученные из газет и социальных сетей.

Проекты, в которых НБФ является наблюдателем. Между независимыми исследованиями и активным участием библиотеки находится группа проектов, в которых библиотека оказывает дополнительные услуги, но не участвует напрямую. Как правило, библиотека предоставляет информацию для таких проектов, способствует получению новой информации или развитию навыков.

Назовем в качестве примера подпроект FinClariah³, осуществляемый Национальным архивом Финляндии. FinClariah — это исследовательская инфраструктура для социаль-

ных и гуманитарных наук (Social Sciences and Humanities, SSH). Общие цели подпроекта заключаются в значительном улучшении доступа к этой инфраструктуре и использовании ресурсов по всем дисциплинам в области SSH.

В рамках подпроекта университеты-участники продолжают разработку инструментов для распознавания именованных сущностей (личных имен, названий организаций и географических названий) в обширном наборе данных оцифрованных архивных файлов. После реализации нескольких недавних проектов и целенаправленного развития в этой области НБФ обладает опытом в распознавании именованных сущностей. Предполагается, что в качестве наблюдателя НБФ сможет предоставить важную информацию для проекта и извлечь выгоду из связанной с ним деятельности.

Проекты с активным участием НБФ.

Роль библиотеки в проекте очень заметна, когда она активно и непосредственно участвует в работе. Степень участия может варьироваться от выполнения четко определенной задачи во внешнем проекте до внутреннего управления всем проектом. Ожидания библиотеки от проекта и ожидания партнеров от библиотеки могут меняться в зависимости от доли ресурсов проекта, выделяемых ей. NewsEye и Translocalis представляют собой два примера проектов такого типа.

Проект NewsEye⁴ финансировался программой исследований и инноваций «Горизонт 2020» Европейского союза. В проекте, рассчитанном на три с половиной года и завершившемся в январе 2022 г., приняли участие три национальные библиотеки, а также исследовательские группы по гуманитарным и социальным наукам, по компьютерным технологиям из пяти университетов. Основной целью проекта была разработка методов и инструментов для эффективного изучения и эксплуатации оцифрованных газет; в нем использовались новые технологии и подходы «больших данных», сочетались методы «близкого» и «удаленного чтения» ДН.

Особенно заметным достижением NewsEye для НБФ стало значительное улучшение качества оптического распознавания символов ранних газет, напечатанных готическим шрифтом. В ходе проекта отдельные газеты повторно распознавались с помощью Transkribus. Поскольку результаты оказались положительными, НБФ заказала повторное распознавание всех оцифрованных газет, изданных до 1914 года.

НБФ участвует в проекте Translocalis⁵, которым руководит Центр мастерства в области истории опыта (Centre of Excellence in the History of Experiences, HEX) Академии Финляндии, расположенный в Университете Тампере. Проект собирает и изучает читательские письма, которые были написаны от имени местных сообществ и опубликованы в финских газетах в XIX веке. HEX вручную собрал письма из оцифрованных газетных материалов с помощью предоставленных НБФ инструментов. В рамках проекта создается база данных, содержащая все письма читателей, опубликованные в период с 1775 по 1885 г., а НБФ разрабатывает новые функции в системе Digi для поддержки использования этой базы данных.

Практические проблемы в исследовательских проектах

Проанализируем проблемы, связанные с различными исследовательскими проектами в области ДН, в которых применялись оцифрованные газеты. Наблюдения основаны как на опыте, полученном в ходе практической работы, так и на информации, собранной по результатам трех опросов пользователей: 2018 г. [3], 2019 г. [4] и 2022 г. [5]. Также выделяются дополнительные вопросы, выявленные в результате исследования, выполненного для проекта «Цифровая открытая память» (Digital Open Memory, DAM). Проект проводился Университетом прикладных наук Юго-Восточной Финляндии [6].

При реализации в конвейерной модели первым этапом является выбор аналоговых данных для оцифровки. С точки зрения библиотеки существует ряд конкурирующих требований, которые определяют параметры отбора, итоговый результат и включают в себя состояние материала и его статус в каталоге. Эти требования не всегда соответствуют ожиданиям ученых.

Оцифровка газетных фондов НБФ еще не завершена; некоторые материалы с конца 1940-х и по 2017 г. еще не обработаны. При нынешних темпах оцифровки библиотеке потребуется не менее 10 лет, чтобы заполнить существующие пробелы. Как правило, невозможно отобрать газеты для оцифровки, исходя из потребностей одного конкретного проекта, однако некоторые из недавних подшивок были оцифрованы в сотрудничестве с издательствами или фондами, вовлеченными в специальные проекты. Такие

проекты помогают восполнить пробелы в коллекциях. Однако, если в ходе этих проектов не будет выделено достаточно дополнительных ресурсов, их реализация может замедлить выполнение других мероприятий по оцифровке.

Комплексная модель, основанная на интерпретирующем рабочем процессе, требует открытой процедуры отбора и оценки в библиотеках, однако это трудно реализовать на практике. Например, имеются ограничения в общедоступной информации, обосновывающей крупномасштабную оцифровку газетных фондов НБФ и причины, по которым некоторые подшивки оцифровывались в течение более длительного времени.

Второй этап конвейерной модели включает в себя операции оцифровки: распознавание текста, распознавание расположения и разделение статей. Распознавание текста и в некоторой степени распознавание расположения уже давно являлись частью стандартных процедур НБФ, в то время как разделение статей не применялось из-за отсутствия передовых автоматизированных инструментов.

Как правило, в ходе рабочего процесса накапливаются проблемы с качеством распознавания текста, однако некоторые устойчивые к ошибкам средства могут в какой-то степени их скрыть. Как показал опрос пользователей Digi 2022 г., наиболее востребованная функция — свободный поиск по тексту. Качество распознавания в различных материалах варьируется, и это существенно влияет на релевантность результатов свободного поиска по тексту. Однако, судя по ответам на опрос, пользователи были удовлетворены результатами поиска; в связи с этим нам необходимо рассмотреть, насколько хорошо ученые понимают важность качества распознавания текста.

Разделение статей было реализовано в нескольких недавних проектах, в том числе NewsEye. Пользовательские интерфейсы включали некоторые продвинутое функции, ожидалось, что успешное применение таких инструментов повысит внимание исследователей к разделению статей. Однако интерес ученых к этой области все еще относительно низок, результаты последнего опроса показывают, что повышения интереса не произошло.

Организация — это третий этап, который включает в себя, в частности, регулирование вопросов авторских прав. Хотя в отношении старых материалов проблема авторских прав не встает, возможно ограничение по работе с публикациями, которым менее 150 лет. В опросах и

в отзывах службы поддержки часто появлялись просьбы увеличить количество материалов в открытом доступе.

Исследователи не всегда разбираются во всех аспектах законодательства об авторском праве и ограничениях, которые оно налагает на библиотеки [6]. Данный пробел в знаниях иногда вызывает ненужные трения между библиотекой и учеными. С точки зрения исследователей, срок защиты авторских прав, действующий в течение 70 лет со дня смерти автора, — чрезвычайно длительный. Даже если материал доступен для использования в библиотеке, такой доступ может не удовлетворить потребности исследователей ДН, которым часто требуются электронные инструменты и вычислительные мощности, недоступные в помещениях библиотеки.

Например, участвующие в NewsEye историки хотели изучать газеты 1850—1950 гг. с помощью инструментов, разработанных ими для проекта. Однако из-за ограничений авторского права библиотека не могла передать более новые материалы, и ученые смогли работать только с газетами из фондов НБФ, опубликованными до 1919 года.

Обе модели (и конвейерная, и комплексная) признают, что при оцифровке создаются данные для последующего применения. Чтобы максимально их использовать, ответственная за оцифровку библиотека должна иметь достаточное представление о тех инструментах, которые потребуются на будущих этапах. Однако на практике бывает трудно найти информацию о соответствующих инструментах.

В моделях также указано, что библиотеки отвечают за составление адекватных описаний материалов. В процессе оцифровки в НБФ сгенерированные необходимые метаданные прикрепляются к оцифрованным материалам. При этом метаданные часто имеют ограничения и могут не удовлетворять всем потребностям пользователей, например трудно получить информацию об отсутствующих в машиночитаемой форме выпусках.

В настоящей статье рассказывается об обеих моделях междисциплинарного сотрудничества для обозначения общих принципов. Эти модели обеспечивают хорошую основу для описания важных вопросов, встающих перед участниками в проекте сторонами, однако они не охватывают все заслуживающие внимания аспекты. Например, разработка программного обеспечения и ручное аннотирование, как пра-

вило, не учитываются в процессе исследования и поэтому легко упускаются из виду на стадии планирования проекта. Иногда такие проблемы можно решить по мере их возникновения; однако более сложными являются задачи, которые необходимо выполнить до и после завершения проекта.

Недавно НБФ решила уделить особое внимание вопросам, связанным с этапом подготовки проекта. В их числе — поиск специалистов, обладающих достаточным опытом для оценки осуществимости проекта, и обеспечение равного отношения к исследователям. Одна из требующих внимания областей — то, каким образом ученые могут предлагать библиотеке сотрудничество. На сегодняшний день они, как правило, обращаются к сотрудникам библиотеки, которые им уже знакомы; таким образом, заинтересованность или рабочая нагрузка отдельного человека могут оказать значительное влияние на ход инициативы. Следует учитывать, что такой неформальный подход несет риски как для библиотеки, так и для исследователя.

Часто исследователи не имеют полного представления о роли, которую библиотека может сыграть в предлагаемом сотрудничестве. Об этом свидетельствует тот факт, что они часто обращаются в библиотеку только в конце этапа планирования проекта, даже если для его реализации требуются значительные библиотечные ресурсы. С библиотечной точки зрения было бы более выгодно принимать участие в планировании на ранней стадии, чтобы до подачи заявки на финансирование оценить ресурсы, которые потребуются на других этапах проекта.

В конце выполнения проекта необходимо устойчивое управление конечными результатами. Итоги реализации некоторых проектов, например NewsEye, можно относительно легко использовать в библиотеке; однако этот процесс не всегда является простым или даже возможным. Одна из причин состоит в том, что традиционные методы гуманитарных исследований выдают данные, которые могут быть понятны только ученым, участвующим в конкретном проекте. Эта проблема становится особенно значительной, если независимый исследователь связывается с библиотекой в конце проекта и просит ее принять на себя ответственность за сохранение результатов и предоставление к ним доступа для будущего использования.

В то же время недавний опрос показал, что лишь около 20% ученых после завершения своих проектов помещают материалы в хранилище данных [7]. Это указывает на то, что количество результативных данных, предоставляемых библиотекам, по-прежнему ограничено и в целом библиотеки не подготовлены к управлению обширными хранилищами данных.

Попытки НБФ решить проблемы

Для удовлетворения потребностей пользователей НБФ постоянно стремится к совершенствованию работы. Придерживаясь основных принципов своей деятельности⁶ — открытости, обновления, образования и совершенствования, НБФ пытается решить указанные выше проблемы. Какие-то решения могут помочь ряду проектов, некоторые ориентированы на конкретные случаи.

Увеличение объема доступной информации приносит особую пользу независимым проектам. Кроме того, это помогает в работе тем ученым, чьи проекты требуют высокого уровня вовлеченности библиотеки.

НБФ опубликовала Программу оцифровки на 2021–2024 гг.⁷ с обзором материалов, которые будут переведены в электронный формат в ближайшие годы. В документе объясняются критерии отбора коллекций, сам процесс и связанные с этим ресурсы. Кроме того, четко очерчены стратегические цели, включая оцифровку обширных унифицированных комплектов ресурсов, а также поддержку открытой науки и разработку методов и инструментов для ДН.

Когда в пользовательский интерфейс Digi добавляется новая информация, НБФ стремится информировать своих читателей о решениях, лежащих в основе развития существующей оцифрованной коллекции. Не все решения были задокументированы с достаточной точностью; например, может быть информация о том, что какой-то элемент был оцифрован в рамках определенного проекта, но может отсутствовать четкое объяснение, почему этот элемент был выбран.

Поскольку положительные результаты проекта NewsEye оказались обнадеживающими, НБФ обратилась к Transkribus с просьбой провести повторное распознавание старейших газет коллекции. Считается, что более качественные данные смогут поддержать все проекты, в которых используются оцифрованные

газеты. При этом важно предоставить ученым информацию о том, что внесенные изменения могли привести к различиям как в результатах поиска, так и в анализе данных.

Независимые исследователи уже несколько лет имеют возможность загружать оцифрованные газеты в виде пакетов открытых данных. Отзывы показали, что у некоторых возникали трудности с готовыми пакетами в своих проектах. Для обеспечения гибкости при загрузке больших объемов материалов НБФ внедрила инструмент, позволяющий пользователю загружать на свое устройство любую часть материалов из открытого доступа. НБФ также занимается метаданными, которые были недостаточны и не соответствовали запросам всех пользователей; внутренняя группа по описанию данных в настоящее время работает над улучшением метаданных, этот проект должен завершиться в 2022 году.

Разделение статей является целью для будущих проектов, и это уже учтено в плане развития системы Digi. Для проекта по оцифровке нотных изданий в систему была интегрирована возможность применения составных компонентов, а разделение статей, реализованное при оцифровке нескольких журналов, стало доступным через пользовательский интерфейс. Таким образом, основа для этого функционала существует, и автоматизация разделения статей доведена до полезного уровня.

НБФ пытается урегулировать вопросы ограничений, связанных с защитой авторских прав, с помощью лицензирования материалов как для публичного онлайн-использования, так и для ограниченного (со стороны исследователей). Законом об авторском праве Финляндии установлена расширенная коллективная лицензия, посредством которой организация как правообладатель может распространять лицензию на авторов и других правообладателей, не представленных организацией. Это позволило НБФ достичь соглашения с организацией по защите авторских прав Koriosto о предоставлении открытого онлайн-доступа к старейшим оцифрованным газетам и журналам Финляндии, включая все время до конца 1939 года. НБФ также заключила еще одно соглашение, которое предоставляет аутентифицированным пользователям онлайн-доступ к более новым оцифрованным газетам; это соглашение также позволяет осуществлять интеллектуальный анализ данных в научных целях.

Количество защищенных авторским правом материалов, доступных онлайн, неуклонно увеличивается — это приводит к росту числа вопросов об их применении. Пользовательский интерфейс Digi отображает четкую информацию о лицензиях, чтобы исключить нарушение правил и условий по незнанию. Эта информация также направлена на то, чтобы помочь независимым исследователям спланировать управление своими данными, а также повысить осведомленность о защищенных авторским правом материалах, которые могут быть недоступны для использования.

НБФ изучила возможность внедрения модели библиотечной лаборатории в рамках проекта DAM. Хотя лаборатория не была создана, стало очевидно, что многие из ее методов необходимо принять. Например, в связи с тем, что некоторые оцифрованные материалы, такие как самые новые газеты, не включены в лицензионные соглашения, библиотека в тестовом режиме запустила модель резидентуры. В этом случае ученые получают более широкий доступ к оцифрованным материалам внутри библиотеки и, в отличие от обычных локальных пользователей, имеют возможность применять различные инструменты анализа.

НБФ объявила о запуске новой процедуры для разъяснения порядка рассмотрения инициатив, связанных с исследовательским сотрудничеством. Лицам, заинтересованным в совместных проектах, теперь предлагается для связи с библиотекой заполнить стандартизированную форму, а не отправлять электронные письма отдельным сотрудникам библиотеки. Предоставлена возможность ввода однотипных входных данных для проектов на этапе оценки. При этом НБФ старается информировать о сроках обработки инициатив в библиотеке, что должно помочь сократить количество запросов, поступающих в последнюю минуту.

Проект Translocalis является хорошим примером эффективного сотрудничества между научным проектом и библиотекой. Исследовательская группа и НБФ контактировали на этапе подготовки проекта и достигли соглашения относительно интеграции базы данных проекта в систему Digi. Таким образом, была предоставлена информация о типе требуемых данных, и НБФ смогла приступить к созданию необходимого функционала.

Для того чтобы следить за разработками в области исследований, необходим широкомасштабный диалог с учеными и другими партне-

рами. Помимо участия в научных конференциях и других мероприятиях, НБФ организовала несколько форумов для сотрудничества.

Координационная группа по цифровым гуманитарным наукам является одним из трех консультативных советов, созданных НБФ для разработки услуг совместно с научным сообществом. Координационная группа следит за реализацией политики НБФ в области ДН и при необходимости дает свои рекомендации в отношении руководящих принципов этой политики. Группа также оценивает совместимость служб обработки данных НБФ с используемыми ДН методами и, чтобы избежать дублирования, обсуждает вопросы распределения обязанностей с другими организациями-членами; такие обязанности включают, например, услуги по сохранению цифровых данных.

Газетный симпозиум — это платформа для взаимодействия между НБФ и исследовательским сообществом. Симпозиум предоставляет НБФ информацию о текущих проектах и дает возможность обсудить планы НБФ с учеными в менее формальной обстановке, чем координационная группа. В отличие от генеральной конференции, у симпозиума конкретная тематика; это позволяет более подробно рассмотреть вопросы, связанные с исследованиями газет. На симпозиуме также можно получить актуальную информацию об инструментах и методах, применяемых в этой области.

Несколько лет назад НБФ запустила серию практических занятий по данным для исследовательского сообщества и аспирантов. Формат занятий позволяет проводить открытые дискуссии и предоставляет информацию о ресурсах НБФ в области обработки данных. Цель проекта состоит в том, чтобы улучшить понимание процессов библиотеки и ее потенциальной роли в проектах в области ДН. Сессия состоит из двух частей: во-первых, презентация в стиле лекции предоставляет справочную информацию о ресурсах и процессах НБФ в области обработки данных; во-вторых, участники знакомятся с оцифрованными данными НБФ и, таким образом, получают практическое представление об их свойствах и исследовательских возможностях.

Обсуждение

В некоторых случаях проблемы, возникающие в проектах ДН, могут быть решены путем внесения дополнительной информации

или внедрения технических усовершенствований, включая оцифровку данных более высокого качества. В конечном счете важнейшим фактором успеха является эффективная коммуникация между библиотекой и исследователями, как минимум библиотечные услуги должны способствовать использованию фондов учеными.

С. Обербихлер предлагает концепцию рабочего процесса в качестве инструмента для обеспечения лучшей коммуникации и рекомендует внедрять его в начале каждого междисциплинарного проекта [1]. Даже если какой-то конкретный рабочий процесс не будет составлен, эта модель может стать полезной отправной точкой для планирования междисциплинарного сотрудничества в исследовательских проектах: неявные предположения будут выражены четко, произойдет улучшение понимания и мотивирование задач, которые в противном случае могли бы быть упущены из виду.

Такие мероприятия, как практические занятия и Газетный симпозиум, стали хорошим способом начать разговор с исследователями ДН, которые ранее не участвовали в проектах с НБФ. Важность этих обсуждений заключается в том, что они предоставляют обеим сторонам новую информацию и открывают новые возможности для сотрудничества.

Библиотеке следует быть открытой для изучения новых направлений и возможностей, но необходимо позаботиться и о том, чтобы исследователи полностью понимали весь диапазон существующих библиотечных услуг. Хорошо описанные и прозрачные базовые процедуры должны способствовать успеху любого совместного проекта.

Примечания

- ¹ <https://whatisdigitalhumanities.com/>
- ² <https://digi.kansalliskirjasto.fi/>
- ³ <https://www.kielipankki.fi/organization/roadmap/>
- ⁴ <https://www.newseye.eu/>
- ⁵ <https://projects.tuni.fi/translocalis-en/>
- ⁶ <https://urn.fi/URN:NBN:fi-fe2020042822711>
- ⁷ <https://urn.fi/URN:ISBN:978-951-51-7103-0>

Список источников

1. Oberbichler S., Boroş E., Doucet A. et al. Integrated Interdisciplinary Workflows for Research on Historical Newspapers: Perspectives from Humanities Scholars, Computer Scientists, and Librarians //

- Journal of the Association for Information Science and Technology. 2021. Vol. 73, № 2. P. 225–239. DOI: 10.1002/asi.24565.
2. Norrgård S., Helama S. Historical Trends in Spring Ice Breakup for the Aura River in Southwest Finland, AD 1749–2018 // *The Holocene*. 2019. Vol. 29, № 6. P. 953–963. DOI: 10.1177/095968361983142.
3. Pääkkönen T., Lilja J. Hieno palvelu, mutta sisältää lisää — Kansalliskirjasto kyseli Digi — palvelun käyttökokemuksia // *Tietolinja*. 2018. № 2. URL: <http://urn.fi/URN:NBN:fi-fe2018092336401> (дата обращения: 01.11.2022).
4. Pääkkönen T. Internal Report. The National Library of Finland, 2019.
5. Pääkkönen T. Syvään päähän: Digi.kansalliskirjasto.fi-palvelun käyttäjäkyselyn tuloksia // *Tietolinja*. 2022. № 1. URL: <https://urn.fi/URN:NBN:fi-fe2022061546944> (дата обращения: 01.11.2022).
6. Kulttuuriperintöä käyttäjä edellä. Tuloksia Digitaalinen avoin muisti — hankkeesta / ed. by M. Kosonen. Kaakkois-Suomen ammattikorkeakoulu, 2021. 49 p. URL: <https://urn.fi/URN:ISBN:978-952-344-357-0> (дата обращения: 01.11.2022).
7. Näpärä L., Lilja J. Kansalliskirjaston aineistojen jatkokäyttö tutkijakyselyssä // *Informaatiotutkimus*. 2021. Vol. 40, № 1. P. 27–62. DOI: 10.23978/inf.99425.

Перевод:

Артём Евгеньевич Зуев,

*Российская государственная библиотека,
международный отдел,
главный библиотекарь
Воздвиженка ул., д. 3/5,
Москва, 119019, Россия
E-mail: mbs@rsl.ru*

Digitized Finnish Newspapers in Digital Humanities Research Projects: Challenges and Solutions from the Library Perspective

Juha Rautiainen,

National Library of Finland, Saimaankatu, 6, Mikkeli, 50100, Finland
E-mail: juha.rautiainen@helsinki.fi

Abstract. Libraries have a direct influence on digital humanities (DH) research projects that utilize digitized newspapers from their collections. However, a library's level of participation in a project will vary and may range from resource provider to active collaborator. Libraries must manage the different expectations of each project; thus, a variety of services are needed.

The National Library of Finland continuously aims to improve its services and provide researchers with broader access to digitized collections. The library's DH research team has extensive experience in research projects, and they actively support the work of DH researchers. In this paper, I present an overview of the experiences of the DH research team. I focus on projects that have used digitized historical newspapers and journals, and I outline a range of potential problems and suggest possible solutions.

Key words: research, newspapers, digitization, digital humanities, library and information service.

Citation: 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.

References

1. Oberbichler S., Boroş E., Doucet A. et al. Integrated Interdisciplinary Workflows for Research on Historical Newspapers: Perspectives from Humanities Scholars, Computer Scientists, and Librarians, *Journal of the Association for Information Science and Technology*, 2021, vol. 73, no. 2, pp. 225–239. DOI: 10.1002/asi.24565.
2. Norrgård S., Helama S. Historical Trends in Spring Ice Breakup for the Aura River in Southwest Finland, AD 1749–2018, *The Holocene*, 2019, vol. 29, no. 6, pp. 953–963. DOI: 10.1177/095968361983142.
3. Pääkkönen T., Lilja J. Hieno palvelu, mutta sisältöä lisää – Kansalliskirjasto kyseli Digi – palvelun käyttökokemuksia, *Tietolinja*, 2018, no. 2. Available at: <http://urn.fi/URN:NBN:fi-fe2018092336401> (accessed 01.11.2022).
4. Pääkkönen T. *Internal Report. The National Library of Finland*, 2019.
5. Pääkkönen T. Syvään päähän: Digi.kansalliskirjasto.fi-palvelun käyttäjäkyselyn tuloksia, *Tietolinja*, 2022, no. 1. Available at: <https://urn.fi/URN:NBN:fi-fe2022061546944> (accessed 01.11.2022).
6. Kosonen M. (ed.) *Kulttuuriperintöä käyttäjä edellä. Tuloksia Digitaalinen avoin muisti – hankkeesta*, 2021. Available at: <https://urn.fi/URN:ISBN:978-952-344-357-0> (accessed 01.11.2022).
7. Näpärä L., Lilja J. Kansalliskirjaston aineistojen jatkokäyttö tutkijakyselyssä, *Informaatiotutkimus*, 2021, vol. 40, no. 1, pp. 27–62. DOI: 10.23978/inf.99425.

Translated into Russian by:

Artem E. Zuev,
Russian State Library,
3/5 Vozdvizhenka Str.,
Moscow, 119019, Russia
E-mail: mbs@rsl.ru

Новая книга



Исаченко Т.А. Святая гора Афон в сказаниях, воспоминаниях и зарисовках русских паломников XVI–XIX веков / Т.А. Исаченко ; Российская гос. б-ка. Москва : Пашков дом, 2023. 443, [1] с. : ил.

В новой работе доктора филологических наук, лауреата Макариевской премии, главного научного сотрудника Центра по исследованию проблем развития библиотек в информационном обществе Российской государственной библиотеки Татьяны Александровны Исаченко рассматриваются источники, вывезенные с Афона на Русь или составленные на Афоне, переписанные в русских монастырях и записанные по «расспросным речам» со слов афонских монахов. Особое внимание уделено автографам XVII в. — чудовского иеродиакона Дамаскина и его «сказанию», «Проскинитарию» Иоанна Комнина Моливда, редкому автографу преподобного Паисия Величковского, обнаруженному в собрании Е.В. Барсова, афонским текстам в библиотеках наместников Свято-Троице-Сергиевой лавры Антония (Медведева) и Леонида (Кавелина). Значительный ряд документов публикуется впервые.

Приобрести книгу:

Книжный магазин Российской государственной библиотеки,
Москва, ул. Воздвиженка, д. 3/5, 3-й подъезд
Тел.: +7 (495) 695-59-53, +7 (499) 557-04-70, доб. 26-46
E-mail: Pashkov_Dom@rsl.ru, sale.pashkov_dom@rsl.ru
Сайт: www.rsl.ru/pashkovdom