

Веб-архивирование в формировании фондов центральных библиотек субъектов РФ: результаты опроса

Веб-архивирование в России находится на начальном этапе развития. Первые попытки реализации проектов в этой области датируются 2000 годом, когда группой экспертов под руководством Министерства РФ по связи и информатизации был разработан проект «Концепции государственного регулирования негосударственными информационными ресурсами России». Целью данного проекта являлось «создание системы архивирования социально значимых ресурсов Интернета» (п. 3.4) [40]¹, но никаких практических последствий это не имело. В действующей редакции Государственной программы «Информационное общество» данная цель также не упоминается [29].²

Выявленные немногочисленные проекты веб-архивирования в российских библиотеках являются единичными, а сами веб-архивы создаются с использованием различного ПО и собственных методических подходов.

Среди крупных национальных библиотек РФ активно веб-архивированием занимается Президентская библиотека им. Б. Н. Ельцина. Профиль комплектования библиотеки содержит «Раздел V. Фонд сетевых электронных документов», где указаны следующие пункты:

- Определение единицы фонда.
- Типология сетевых электронных документов.
- Срок хранения документов в ФСЭД.
- Предметно-тематическое ядро ФСЭД.
- Критерии отбора документов ФСЭД.

¹ Концепция государственного регулирования негосударственных информационных ресурсов России // Информрегистр: сайт. – URL : <http://www.inforeg.ru/library/item/138> (дата обращения: 23.01.2022).

² Информационное общество : государственная целевая программа / М-во цифрового развития, связи и массовых коммуникаций. – URL : <https://digital.gov.ru/ru/activity/programs/1/> (дата обращения: 02.07.2022).

Среди республик России большое внимание развитию электронных ресурсов уделяет Республика Саха (Якутия). Национальная библиотека Республики не только осуществляет издание неопубликованных документов на условиях лицензионного договора с авторами, но и развивает краеведческий проект «Yakutia.online», направленный на сохранение краеведческих сетевых документов. Цель проекта заключается в электронной каталогизации и создании веб-хранилища электронных сетевых ресурсов (составная часть) в Национальной библиотеке Республики Саха (Якутия) в целях сохранения ценного краеведческого контента.

В целях выявления общей картины развития веб-архивирования в центральных библиотеках субъектов Российской Федерации было проведено анкетирование, которое позволило выявить современное состояние веб-архивирования на уровне центральных библиотек субъектов Российской Федерации. Целью анкетирования является подготовка системных предложений по применению технологии веб-архивирования в практике формирования библиотечных фондов. Анкетирование осуществлялось онлайн на платформе Яндекс. Формы (<https://forms.yandex.ru/cloud/669f6b5773cee70ee73f8c13/>). Вопросы анкеты были согласованы с сотрудниками сектора изучения библиотечных фондов РНБ.

Всего ссылка на анкету была разослана в 85 субъектов Российской Федерации. В опросе участвовала 31 Центральная библиотека региона России (34%). Значительная часть регионов проигнорировала приглашение к опросу в связи с чем был сделан вывод о низком уровне готовности к обсуждению предлагаемой темы. Несмотря на относительно небольшой отклик Центральные библиотеки субъектов РФ, полученный результат позволил выявить репрезентативную картину состояния веб-архивирования в регионах России.

В задачи опроса входила:

- оценка общей ситуации применения технологии веб-архивирования электронных сетевых документов в формировании библиотечных фондов центральных универсальных библиотек субъектов Российской Федерации;
- оценка уровня информированности специалистов о возможности архивирования сетевых документов для включения их архивированных копий в фонд библиотеки;
- оценка возможностей реформирования системы федерального и регионального законодательства об обязательном экземпляре для сбора и хранения электронных документов;
- оценка отражения процессов создания веб-архивов во внутренних нормативных документах библиотеки;
- сбор информации о существующих практиках в сфере сохранения сетевых документов в библиотеках РФ;
- определение наиболее актуальной для регионов тематики выборочного веб-архивирования;
- выявление проблем, возникающих с правообладателями при сохранении сетевых документов, и способов их решения;
- оценка перспектив применения технологий с открытым исходным кодом для архивирования сетевых документов;
- разработка предложений по применению технологии веб-архивирования в формировании библиотечных фондов.

Одной из задач проводимого опроса являлась оценка необходимости архивирования сетевых документов в электронных фондах библиотеки. Для этого был задан вопрос, в результате которого выявлялось субъективное мнение специалиста о том, нужен ли веб-архив в фонде библиотеки региона. Большинство специалистов поддержали идею о создании веб-архивов в библиотеках регионов (84%).

В качестве аргументов специалистами региональных библиотек были предложены следующие:

- Архивирование позволяет сохранить важные сетевые электронные документы, подверженные риску утраты, для будущих поколений (Кировская область, Пермский край, Забайкальский край, Смоленская область, Воронежская область, Ярославская область);

- Сохранение редкой краеведческой информации (Архангельская область, Ставропольский край, Самарская область);

- Мемориальная функция библиотеки должна включать сохранение электронных документов, в том числе сетевых документов (Республика Коми);

- Некоторые электронные документы могут иметь юридическую или отчетную ценность (например, отчеты, договоры, результаты мероприятий). Архивирование обеспечивает соблюдение законодательных требований (Кировская область);

- Историческая ценность. Сетевые документы, такие как публикации на сайте или в социальных сетях, могут со временем стать источником исторической информации о деятельности библиотеки и ее развитии (Кировская область).

Были также выдвинуты аргументы против реализации веб-архивирования в отечественных библиотеках. В основном они включали в себя проблемы финансирования и недостаточную разработанность правовых норм в части сетевых документов (Удмуртская республика, Иркутская область). По мнению специалиста Иркутской областной государственной универсальной научной библиотеки им. И. И. Молчанова-Сибирского, «у библиотек нет ни технических, ни кадровых ресурсов. Особенно если мы говорим о выборочном архивировании по тематическим профилям и в привязке к географии. То есть, если ее и в рамках Министерства культуры, то только на базе федерального учреждения, которое может себе позволить тотальное архивирование доменов ru/rf/su. Любое выборочное архивирование требует слишком больших затрат на поиск, оценку и отбор объектов».

Проведённый опрос позволил оценить общую ситуацию реализации веб-архивирования электронных сетевых документов в сфере формирования библиотечных фондов центральных универсальных библиотек субъектов Российской Федерации как находящуюся в начальной фазе своего развития. Большинство опрошенных специалистов были заинтересованы в сохранении сетевых электронных фондов библиотек и поддерживали идею сохранения цифрового наследия в библиотечных фондах.

Несмотря на то, что некоторые респонденты продемонстрировали достаточный уровень информированности о возможности архивирования сетевых документов для участия в опросе и предложений по техническим и правовым аспектам, большинство специалистов библиотек мало знакомо с веб-архивированием как отраслью библиотечной деятельности и особенностями выборочного веб-архивирования, что подтверждается низким показателем участия в опросе по регионам РФ.

Специалисты ЦБ регионов РФ выразили необходимость реформирования системы федерального и регионального законодательства об обязательном экземпляре для сбора и хранения электронных документов. Была выявлена проблематика внедрения электронного обязательного экземпляра в систему нормативно-правовых актов разных уровней, касающаяся определения границ обязательного электронного экземпляра, организации региональных репозиторий для хранения цифровых ОЭ, определения правообладателей комплексных электронных документов.

Большинство специалистов поддержали тезис о необходимости отражения процессов создания веб-архивов во внутренних нормативных документах библиотеки при наличии федеральной и региональной нормативно-правовой базы для веб-архивирования и разработанности методологического подхода к архивированию сетевых документов.

В результате опроса было выявлено, что существующие практики в сфере сохранения сетевых документов в библиотеках РФ реализуются в двух направлениях:

- Архивирование региональной периодики в рамках создания краеведческих порталов и коллекций.

- Архивирование собственного контента библиотек в целях предотвращения утраты информации в результате сбоев ПО и хакерских атак.

Наиболее актуальными для регионов по видам ресурсов (периодика и копии печатных местных изданий) и тематике выборочного веб-архивирования стали история и культурные особенности краевых народов региона, местная топонимия. В целом выделяется краеведческое направление с учётом отражения роли региона в крупных государственных процессах и событиях.

Основной проблемой правового характера при сохранении сетевых документов является отсутствие нормативно-правовой базы для архивирования сетевых документов. Поскольку большинство опрошенных библиотек не занимается веб-архивированием, то для решения правовых проблем применяются в основном прямые договоры с правообладателями, когда речь идёт о региональной периодике.

В ходе опроса была выявлена осведомленность большей части в опрошенных о существовании технологий с открытым исходным кодом для архивирования сетевых документов, таких как Heritrix и Htttrack. Согласно отзывам специалистов, для освоения подобных технологий требуются дополнительные компетенции специалистов. Кроме того, было заявлено о недостаточных возможностях бесплатных решений для веб-архивирования.

Об использовании приложений компании Webrecorder было заявлено специалистами Архангельской областной центральной научной библиотеки. Данное приложение позволяет архивировать комплексный сетевой документ (веб-сайт) в формате WARC. Помимо возможности веб-архивирования, компания также предлагает бесплатный плеер для архивированных страниц и браузерное расширение для ручного сохранения веб-страниц с возможностью загрузки на сервер и последующего воспроизведения. Данная технология наиболее удобна в рамках текущего исследования, поскольку отвечает задачам

экспертного глубокого харвестинга, выражающегося в оперативном формировании коллекций архивированных сетевых документов.

Опыт специалистов библиотек, представленный в результатах опроса позволил сформулировать следующие рекомендации для выбора ПО для архивирования сетевых документов:

- использовать бесплатное ПО с открытым исходным кодом для снижения финансовой нагрузки и предотвращения правовых издержек использования кода;
- выбрать ПО с максимально доступным для пользователя интерфейсом, который не требует изучения языков программирования (как Heritrix, для работы с которым требуется установка среды Java).

Наиболее актуальными проблемами, стоящими на пути развития веб-архивирования, по мнению респондентов опроса являются проблемы недостаточного финансирования, материально-технического обеспечения, отсутствия конкретной нормативно-правовой базы и методологии, разработанной с учётом нынешних реалий.

По данным опроса, для методики селективного веб-архивирования требуется доступность её реализации, которая основана на сокращении предполагаемых объёмов формируемой коллекции за счёт экспертного отбора архивируемых материалов, использования доступного в финансовом и исполнительском аспектах ПО, а также чётких критериях выбора тематики и отбора документов для веб-архивирования.

В целом процесс внедрения цифрового обязательного экземпляра с конкретизацией аспектов работы с сетевыми документами может занять длительный срок, что значительно усилит отставание отечественных библиотек в сфере архивирования сетевых документов. Цифровой обязательный экземпляр является неременным условием развития библиотечно-информационной отрасли в плане формирования электронных ресурсов и интеграции библиотеки в сферу глобальных проектов сохранения мирового культурного наследия и информационного обмена. Тем не менее,

для реализации выборочного веб–архивирования в отечественных библиотек и снижения утраты ценных сетевых документов целесообразно использовать текущее законодательство и формировать коллекции сетевых документов, посредством выборочного архивирования.

В результате анализа источников, практики работы библиотек в области веб-архивирования и результатов проведенного опроса специалистов Центральных библиотек субъектов РФ был сделан вывод о том, что веб–архивирование в библиотеках России разных типов и уровней находится на начальном этапе. Основными проблемами, сдерживающими развитие этого процесса связано с отсутствием нормативно-правовой базы, в которую необходимо было бы включить:

- представления о процессах веб-архивирования и специфику документов сетевого происхождения и особенности их учета;

- особенности реализации авторского права в процессе веб-архивирования, включив в нормативные акты «механизм отказа», который позволяет библиотекам копировать и сохранять в своих фондах контент без запроса согласия правообладателя, но удалять его по его требованию;

- особенности учета сиротских произведений, опубликованных в Сети (их эфемерность, сложность атрибуции, вызванная нестандартным оформлением и др.), а также произведений, созданных или опубликованных непрофессиональными и просто «случайными» авторами, делающими практически невозможным их лицензирование;

- обоснование сохранения цифрового культурного наследия в библиотеках как институтах сохранения социальной памяти;

- зафиксировать единые для библиотек принципы учёта сетевых документов, не являющихся электронными версиями печатных изданий, в фонде библиотеки, а также формирования раздела в профиле комплектования.

Сложной проблемой, которую необходимо решать создателям веб-архивов является взаимодействие с пользователями при формировании коллекций архивированных электронных документов, которые могут сами

внести вклад в библиотечное веб–архивирование, предлагая свои электронные документы.

Российские научно-теоретические исследования в области веб–архивирования в основном посвящены обзору деятельности зарубежных библиотек и изучению зарубежных практик по сохранению такого вида документов, а также отдельным узким аспектам создания веб-архивов.

Выявленные немногочисленные проекты веб-архивирования в российских библиотеках являются единичными, а сами веб-архивы создаются с использованием различного ПО и собственных методических подходов.

Несмотря на то, что многие специалисты библиотечно-информационной сферы активно поддерживают необходимость внедрения веб-архивирования в деятельность региональных библиотек, оказалось, имеются и аргументы против, связанные с потенциально высокими затратами на организацию региональных репозиториев, правовыми проблемами архивирования сетевых документов, недостаточной разработанностью отечественного ПО для решения подобных вопросов и проблемами увеличения нагрузки на персонал библиотек.

Большинством региональных библиотек архивируются документы краеведческой тематики, а также местная периодика (как в печатном, так и в электронном формате). Библиотеки регионов РФ делают шаги в архивировании периодики, которая имеет краеведческую ценность, когда она уже не существует в печатном виде на основе прямых договоров с правообладателями.

Проведённый опрос позволил оценить общую ситуацию реализации веб-архивирования электронных сетевых документов в сфере формирования библиотечных фондов центральных универсальных библиотек субъектов Российской Федерации как находящуюся в начальной фазе своего развития. Большинство опрошенных специалистов были заинтересованы в сохранении сетевых электронных фондов библиотек и поддерживали идею сохранения цифрового наследия в библиотечных фондах.

Специалисты ЦБ регионов РФ выразили необходимость реформирования системы федерального и регионального законодательства об обязательном экземпляре для сбора и хранения электронных документов. Была выявлена проблематика внедрения электронного обязательного экземпляра в систему нормативно-правовых актов разных уровней, касающаяся определения границ обязательного электронного экземпляра, организации региональных репозиторий для хранения цифровых ОЭ, определения правообладателей комплексных электронных документов.

Наиболее актуальной для регионов тематикой выборочного веб-архивирования стали история и культурные особенности краевых народов региона, местная топонимия. В целом выделяется краеведческое направление с учётом отражения роли региона в крупных государственных процессах и событиях.

По данным опроса, для применения единой методики селективного веб-архивирования требуется доступность её реализации, которая основана на сокращении предполагаемых объёмов формируемой коллекции за счёт экспертного отбора архивируемых материалов, использовании доступного в финансовом и исполнительском аспектах ПО, а также чётких критериях выбора тематики и отбора документов для веб-архивирования.

Наиболее актуальными проблемами, стоящими на пути развития веб-архивирования, по мнению респондентов являются проблемы недостаточного финансирования, материально-технического обеспечения, отсутствия конкретной нормативно-правовой базы и методологии, разработанной с учётом нынешних реалий.

Основной проблемой правового характера при сохранении сетевых документов является отсутствие нормативно-правовой базы для архивирования сетевых документов. Поскольку большинство опрошенных библиотек не занимается веб-архивированием, то для решения правовых проблем применяются в основном прямые договоры с правообладателями, когда речь идёт о региональной периодике.

Большинство специалистов поддержали тезис о необходимости отражения процессов создания веб-архивов во внутренних нормативных документах библиотеки при наличии федеральной и региональной нормативно-правовой базы для веб-архивирования и разработанности методологического подхода к архивированию сетевых документов.

Опыт специалистов библиотек, представленный в результатах опроса позволил сформулировать следующие рекомендации для выбора ПО для архивирования сетевых документов:

- использовать бесплатное ПО с открытым исходным кодом для снижения финансовой нагрузки и предотвращения правовых издержек использования кода;

- выбрать ПО с максимально доступным для пользователя интерфейсом, который не требует изучения языков программирования (как Heritrix, для работы с которым требуется установка среды Java).