

## Проблемы отечественного и зарубежного веб-архивирования в библиотеках. Веб-архивирование как область деятельности

**А. А. Смирнов**

*Российская национальная библиотека,  
Санкт-Петербург, Российская Федерация,  
artjomvajda@gmail.com*

**Аннотация.** Статья посвящена актуальным проблемам современного отечественного и зарубежного веб-архивирования. В ходе исследования проводились поиск и анализ международных стандартов и других значимых документов, изучалась контекстная среда (интернациональные проекты и организации), оценивалась российская законодательная база, касающаяся веб-архивирования. Был использован метод индуктивного информационного анализа, в рамках которого рассмотрены актуальные публикации по теме, проектные инициативы, сайты ведущих научных и практических организаций. В статье описаны международные стандарты в области веб-архивирования данных, определены основные проблемы процесса с технической точки зрения, предложены современные виды программного обеспечения для копирования веб-документов. Представлены также возможные пути развития веб-архивирования в свете сохранения личных страниц пользователей: отражение персональных цифровых архивов в веб-коллекциях, посвящённых культуре региона.

Состояние отечественного веб-архивирования рассмотрено на примере Президентской библиотеки им. Б. Н. Ельцина, деятельность которой направлена на сохранение ресурсов, связанных с государственной властью: сайтов правительственных учреждений и структур. Данная группа веб-документов может тиражироваться и распространяться без каких-либо ограничений, поскольку содержит информацию об органах исполнительной власти. Предложен селективный (выборочный) подход к веб-архивированию: копирование и создание коллекций веб-документов, посвящённых социально значимым событиям.

**Ключевые слова:** веб-архивирование, краулер, семантический веб, метаданные, Heritrix, национальные библиотеки, харвестинг, РНБ, электронный обязательный экземпляр, web-archiving, crawler, web-harvesting

**Для цитирования:** Смирнов А. А. Проблемы отечественного и зарубежного веб-архивирования в библиотеках. Веб-архивирование как область деятельности / А. А. Смирнов // Научные и технические библиотеки. 2022. № 12. С. 104–123. <https://doi.org/10.33186/1027-3689-2022-12-104-123>

**Благодарности:** автор выражает благодарность и глубокую признательность своему научному руководителю Т. В. Захарчук, доктору пед. наук, профессору СПбГИК, и И. В. Эйдемиллер, заведующей сектором изучения фондов научно-методического отдела РНБ, за советы и ценные замечания при работе над данной статьёй.

UDC [005.921.1:004.63]:025

<https://doi.org/10.33186/1027-3689-2022-12-104-123>

## The problems of national and foreign web-archiving in libraries. Web-archiving as a functional area

**Artyom A. Smirnov**

*National Library of Russia, St. Petersburg, Russian Federation,  
artjomvajda@gmail.com*

**Abstract.** The author examines the current problems of web-archiving practice in this country and abroad. The international standards and other important documents were retrieved and analyzed; the contextual environment (international projects and organizations) were explored, related Russian legislation is evaluated. The method of inductive information analysis was applied to analyze current publications on the subject, projects and initiatives, and websites of leading research and service organizations. The author analyses international standards in web-archiving, defines key technological problems, and discusses modern software products for copying web-documents. He also analyses the trends in web-archiving in the aspect of saving personal user accounts: allocating personal digital archives in the regional cultural web-collections.

The current status of national web-archiving is analyzed as the case study of B. Yeltsin Presidential Library which mission is to preserve resources related to the powers of the state: websites of governmental organizations and agencies. This group of web-documents can be replicated without any constraints as it comprises information on executive authorities. The author suggests a selective approach toward web-archiving, i. e. replication and building collections of web-documents on socially important events.

**Keywords:** web-archiving, crawler, semantic Web, metadata, Heritrix, national libraries, web-harvesting, National Library of Russia, digital mandatory copy

**Cite:** Smirnov A. A. The problems of national and foreign web-archiving in libraries. Web-archiving as a functional area / A. A. Smirnov // Scientific and technical libraries. 2022. No. 12. P. 104–123. <https://doi.org/10.33186/1027-3689-2022-12-104-123>

**Acknowledgements:** the author thanks and expresses profound gratitude to his thesis adviser Tatiana V. Zakharchuk, Dr. Sc. in Pedagogy, Professor of St. Petersburg State Institute of Culture, and Irina V. Eidemiller, Head of Collection Studies Sector of the Research and Instruction Department of the National Library of Russia, for their advice and insightful comments on this article.

Международный консорциум по архивированию Интернета (ИИРС) определяет веб-архивирование как совокупность процессов сбора частей Всемирной паутины, сохранения коллекций в архивном формате и последующего предоставления архивов для доступа и использования [1]. В сфере мирового регулирования подобного рода процессов на данный момент используется международный стандарт ISO 14721:2012 (Space data and information transfer systems – Open archival information system (OAIS) – Reference model). Этот стандарт регулярно пересматривается и обновляется. Согласно техническому отчёту ISO/TR 14873:2013, веб-архивирование относится к действиям по выбору, сбору, хранению и управлению доступом к исполняемым копиям интернет-ресурсов за определённый промежуток времени [2].

В зависимости от масштаба и цели сбора различают два основных способа веб-архивирования: систематический сплошной автоматиче-

ский сбор и архивирование (харвестинг) и избирательное архивирование (глубокий, или выборочный харвестинг). *Харвестинг* национального домена предназначен для получения содержательно-структурной копии всего домена (или его части). Данный подход характерен для национальных библиотек Ирландии, Хорватии, Финляндии, Австралии, Канады, Китая, Бельгии, Германии, Израиля и других стран.

Для проведения харвестинга, как правило, используется открытое программное обеспечение. Наиболее распространённым в национальных библиотеках и других организациях, занимающихся веб-архивированием, является открытое ПО Heritrix (табл. 3). Веб-краулер компании IA Heritrix делает «моментальный снимок» (англ. snapshot) или копию сайта (файл в формате WARC). Данный файл может содержать контент всех уровней: от сайта целиком, как комплексного документа, связанного гипертекстовыми отношениями, до отдельных объектов, размещённых на нём. Масштаб сбора доменов делает невозможным проведение любой экспертной оценки, которая является распространённым методом обеспечения качества при выборочном харвестинге.

*Выборочный харвестинг* выполняется в гораздо меньших масштабах, он более целенаправлен. Отбор производится исходя из определённых критериев: в центре внимания может быть событие, имеющее особое социально-культурное или историческое значение для страны, тема, персоналия, материалы одного формата (например, аудио-, видеофайлы), или в основе – соглашение с владельцами контента. Ключевое различие между двумя стратегиями заключается в уровне контроля качества, когда собранные веб-сайты оцениваются на предмет соответствия заранее определённым стандартам.

В зависимости от стратегической важности, имеющихся средств, а иногда и требований законодательства, были предприняты разнообразные подходы к архивированию интернет-ресурсов: от записи отдельных веб-страниц до копирования целых доменов верхнего уровня.

Основная цель веб-архивирования состоит в том, чтобы навсегда сохранить записи об Интернете, максимально приближенные к их первоначальной форме, для различных научных, исследовательских, образовательных, профессиональных и других целей.

## **Стандартизация и координация процессов создания веб-архивов**

За рубежом в большинстве случаев веб-архивирование национального домена осуществляется по единым международным стандартам с использованием открытого программного обеспечения Heritrix в WARC-формате. С организационной точки зрения процессы находятся на разных уровнях зрелости. В то время как в одних структурах веб-архивирование стало обычным делом, другие только составили экспериментальные программы для изучения темы.

Ежегодно Международный консорциум сохранения Интернета проводит конференцию Web Archiving Conference (WAC), на которой звучат доклады по разработке технологий для общего и избирательно-го веб-архивирования, а также по проблемам национального законодательства по авторскому праву в цифровой среде.

Среди основных проблем, которые стоят перед национальными библиотеками, формирующими веб-архивы, на первом месте стоит сложность согласования копирования и сохранения данных с правообладателями. По данным консорциума, частные организации предпочитают не отвечать на запросы библиотек о копировании интернет-ресурсов [3].

К наиболее известным организациям, занимающимся большими объёмами архивируемых веб-данных, относится некоммерческая организация Internet Archive (Архив Интернета), основанная в 1996 г. Одноимённый проект аккумулирует копии веб-страниц, графические материалы, видео-, аудиозаписи и программное обеспечение, гарантируя долгосрочное архивирование собранного материала и бесплатный доступ к своим БД для широкой публики. Размер архива на 2019 г. – более 45 петабайт; еженедельно добавляется около 20 терабайт. Основной веб-сервис архива – The Wayback Machine. Содержание веб-страниц фиксируется с определённым временным промежутком с помощью бота. Таким образом, можно посмотреть, как выглядела та или иная страница раньше, даже если она больше не существует по старому адресу [Там же].

Учреждения, занимающиеся веб-архивированием, рассматривают его как продолжение миссии по сохранению национального культурного наследия. В национальных библиотеках разработки в области веб-архивирования ведутся с 1994 г., начиная с пилотного проекта EPPP (Electronic Publication Pilot Project) Национальной библиотеки Канады.

С 1996 г. веб-ресурсы входят в состав объектов обязательного экземпляра многих стран. Особенно активно этот процесс стал развиваться в последние десятилетия (2000–2020 гг.).

В качестве обязательного экземпляра могут выступать: веб-сайты, отдельные страницы и объекты, размещённые на страницах веб-сайтов (изображения, электронные тексты, мультимедиамаериалы и другие виды контента).

Одной из первых стран, в которых обязательный экземпляр был распространён на все цифровые материалы, в том числе интернет-ресурсы открытого доступа, включая блоги, вики и пр., была Новая Зеландия. В соответствии с принятым в 2003 г. законом Национальная библиотека Новой Зеландии получила право на сбор и сохранение всех электронных публикаций и интернет-ресурсов страны. Это позволило основать Национальный архив цифрового наследия (NDNA) для обеспечения бессрочного хранения цифровой информации о Новой Зеландии (табл. 1).

Национальная стратегия, которая лежит в основе NDNA, не делает различий между контентом, созданным авторизованными организациями, и контентом, созданным гражданами. Хранящиеся в NDNA цифровые материалы поступают в НБ из четырёх источников: обязательные экземпляры, интернет-поиск, безвозмездно предоставленные и оцифрованные материалы [4]. Подобные законодательные инициативы в области внедрения принципа обязательного экземпляра в веб-среду были реализованы также в Великобритании, Германии, Норвегии, Австралии, Литве, Эстонии, Японии и других странах (табл. 2).

Таблица 1

**Нормативно-правовое регулирование харвестинга  
веб-документов в странах Европы и Азии  
(согласно данным Международного консорциума по веб-архивированию)**

Удалённый доступ	Веб-архивирование регулируется законодательно. Обязательный электронный экземпляр			Ожидается расширение существующего законодательства	Нет законодательства о веб-архивировании
	Комбинированная политика	Доступ только в помещении библиотеки	Разрешён только сбор		
Словения	Австралия	Австрия	Швеция	Канада (Квебек)	Китай
Норвегия	Новая Зеландия	Великобритания		Сингапур	Чехия
Исландия	Япония	Испания			США
Хорватия		Франция			Польша
		Финляндия			Швейцария
		Эстония			Россия
		Дания			

**Международные стандарты и другие документы  
по веб-архивированию**

Название	Описание	Электронный адрес
ISO 14721:2012 «Системы передачи данных и информации о космическом пространстве. Открытая архивная информационная система. Эталонная модель» (Space data and information transfer systems – Open archival information system (OAIS) – Reference model)	В данном стандарте отражены основные аспекты архивирования Больших данных, разработаны наиболее часто используемые открытые стандарты метаданных: Дублинское ядро, EAD, METS, словарь метаданных для обеспечения долговременного архивного хранения электронных документов – PREMIS (он специально разработан для описания событий/действий с течением архивного хранения).	Вторая редакция стандарта доступна как «сиреневая книга» CCSDS 650.0-M-2 Reference Model for an Open Archival Information System (OAIS), см. <a href="https://public.ccsds.org/pubs/650x0m2.pdf">https://public.ccsds.org/pubs/650x0m2.pdf</a>
ISO 16363:2012 Audit and certification of trustworthy digital repositories (TDR)	Устанавливает исчерпывающие показатели того, что должен делать архив, на основе OAIS.	Доступен как «сиреневая книга»: <a href="https://public.ccsds.org/pubs/652x0m1.pdf">https://public.ccsds.org/pubs/652x0m1.pdf</a>
ISO 28500:2017 Information and documentation – WARC file format	В данном стандарте отражены области применения формата WARC, рассмотрены способы дедупликации файлов в электронных архивах, возможности сохранения метаданных, протоколы сжатия файлов и т. д.	<a href="https://www.iso.org/standard/68004.html">https://www.iso.org/standard/68004.html</a>
ISO/TR 14873:2013(en) Information and documentation – Statistics and quality issues for web archiving	Этот технический отчет определяет статистику, условия и критерии качества для веб-архивирования. Он учитывает потребности и практику широкого круга организаций, таких как библиотеки, архивы, музеи, исследовательские центры и фонды наследия.	<a href="https://www.iso.org/standard/55211.html">https://www.iso.org/standard/55211.html</a>



Название	Описание	Электронный адрес
ISO 16919:2014 Space data and information transfer systems – Requirements for bodies providing audit and certification of candidate trustworthy digital repositories	Стандарт предназначен в первую очередь для тех, кто создаёт организацию, выполняющую аудит и сертификацию цифровых репозиториев, и управляет ею. Он также должен быть полезен для тех, кто работает с цифровыми репозиториями и стремится объективно измерить надёжность своего репозитория.	Доступен как «сиреневая книга»: <a href="https://public.ccsds.org/Pubs/652x1m2.pdf">https://public.ccsds.org/Pubs/652x1m2.pdf</a>

Обязательный экземпляр электронного документа позволяет Британской национальной библиотеке реализовать концепцию «голоса поколения», в рамках которой производится сбор электронного наследия гражданского общества. В сотрудничестве с пятью крупнейшими библиотеками Соединённого Королевства Британская библиотека собирает сайты национального домена и является полноправным членом Всемирного консорциума сохранения Интернета, участвуя в международных проектах по веб-архивированию [5].

Следует особо отметить опыт Национальной библиотеки Австралии. Сегодня она собирает и сохраняет для долгосрочного доступа Австралийский веб-архив, доступный через национальный портал Trove. Коллекция создаётся за счёт выборочного глубокого харвестинга (проект PANDORA), харвестинга веб-сайтов правительства Австралии (AGWA) и ежегодного харвестинга всего домена au.

Согласно исследованию Архива Интернета, национальные библиотеки предпочитают самостоятельно заниматься веб-архивированием. Создание веб-архивов интегрируется в библиотечную практику формирования фондов. Бюджетные сокращения и нехватка ресурсов, которые наблюдались во многих учреждениях, затрудняют разработку необходимых инструментов для веб-архивирования.



Участники исследования отмечают пробелы в содержании веб-архивов. Национальные библиотеки желают сохранять ту часть социальных сетей, которая принадлежит соответствующему национальному домену. Они также хотели бы использовать веб-архивирование в качестве дополнительного инструмента для сбора таких цифровых объектов в Интернете, как электронные книги, электронные журналы, музыка и карты.

Несмотря на то, что национальные веб-архивы в целом мало используются из-за ограничений доступа, многие национальные библиотеки хотят поддерживать использование своих веб-архивов в научных целях, взаимодействуя с учёными для понимания их требований и, в конечном итоге, встраивая коллекции веб-архивов в исследовательский процесс.

Национальные библиотеки прибегают к сотрудничеству с Архивом Интернета для формирования коллекций веб-документов. Ряд национальных библиотек выразил потребность в использовании ПО организации, включая обслуживание, поддержку и новые функции. Среди них, например, Национальная библиотека Ирландии, селективные веб-коллекции которой можно найти на портале Archive-it.org [6].

Одним из проектов Internet Archive является Community Webs, который позволил региональным библиотекам сохранить более 50 терабайт материалов о культурном наследии в Интернете, документирующих жизнь местных жителей, общественных объединений разных видов. Также были созданы открытые образовательные ресурсы, относящиеся к веб-архивированию, цифровому хранению, архивированию сообществ и развитию коллекций. Участники исследовали новые формы партнёрства на местном уровне с помощью государственных программ и краудсорсинга, а их цифровые коллекции использовались учёными и другими специалистами [7].

### **Проблемы сбора веб-документов**

Осуществляя проекты веб-архивирования, национальные библиотеки столкнулись с тем, что при сплошном харвестинге национальных доменов образуются огромные массивы информации, которые плохо поддаются качественному контролю, управлению и обработке. Для обеспечения сохранности таких массивов требуются большие финансовые вложения.

По сведениям Консорциума сохранения Интернета, среди основных трудностей данного процесса выделяется проблема сохранения больших объёмов информации, которая к тому же может со временем исчезать с действующего сетевого адреса или перемещаться на более глубокий уровень веба. При использовании краулера Heritrix возникают проблемы качественного охвата архивных файлов веб-документов, содержащих такие элементы, как:

- динамически масштабируемые изображения, тексты и карты;
- загружаемые файлы мультимедиа, требующие нажатия кнопки «Воспроизвести»;
- меню навигации;
- пагинация на основе JavaScript [8].

Поскольку указанные элементы сегодня являются неизменными атрибутами как сайтов и веб-порталов, так и соцсетей, архивирование которых представляет собой одну из приоритетных задач, то разрабатываются и применяются новые технологии и программы архивирования данных Сети. Среди них технология Brozzler, разработанная в Архиве Интернета, которая полагается на реальный веб-браузер для взаимодействия с веб-контентом до того, как этот контент будет проиндексирован и заархивирован в файлы WARC, а также использует youtube-dl для расширения возможностей захвата мультимедиа (по состоянию на январь 2020 г. и Brozzler, и стандартные обходы используют youtube-dl).

Можно также встретить упоминания набирающей популярность процесса web-packaging, который направлен на пакетное архивирование сетевых объектов для последующего хранения и передачи этих данных. Среди преимуществ, которые пакетное архивирование может дать веб-архивированию, выделяют эффективное сканирование, точное воспроизведение и неотчуждаемый характер электронной копии [9].

Таким образом, продолжают развиваться разработки в области программного обеспечения для формирования веб-архивов, поиска и обработки файлов. При создании новых программных продуктов учитываются такие запросы, как исключение дублирования файлов, включение в архив многоуровневых веб-документов, индексирование содержимого архива и др. (табл. 3).

## Программные средства для веб-архивирования

Название	Разработчики	Описание	Формат файла	Источник
Heritrix	Архив Интернета, США	Heritrix – это поисковый робот, разработанный для веб-архивирования. Он доступен по лицензии бесплатного программного обеспечения и написан на Java.	WARC	<a href="https://github.com/internetarchive/heritrix3">https://github.com/internetarchive/heritrix3</a>
Open Wayback	Лорен Ко, Университет Северного Техаса	OpenWayback – это ключевое программное обеспечение, которое веб-архивы по всему миру используют для «воспроизведения» заархивированных веб-сайтов в браузере пользователя.	ARC или WARC	<a href="https://netpreserve.org">https://netpreserve.org</a>
Apache Solr	Apache Software Foundation	Solr – платформа полнотекстового поиска с открытым исходным кодом, основанная на проекте Apache Lucene. Её основные возможности: полнотекстовый поиск, подсветка результатов, фасетный поиск, динамическая кластеризация, интеграция с БД, обработка документов со сложным форматом.	–	<a href="https://lucene.apache.org/solr/">https://lucene.apache.org/solr/</a>
Web Curator Tool	Национальная библиотека Новой Зеландии, Британская библиотека	Web Curator Tool (WCT) – это приложение для управления рабочим процессом с открытым исходным кодом для выборочного веб-архивирования. Приложение интегрировано с краулером Heritrix и поддерживает ключевые процессы, такие как харвестинг, проверка качества и сбор описательных метаданных [20].	WARC	<a href="http://webcurator-tool.org/">http://webcurator-tool.org/</a>

Название	Разработчики	Описание	Формат файла	Источник
Netarchive Suite	Разрабатывается и поддерживается Королевской датской библиотекой (ранее – Королевской библиотекой Копенгагена и Государственной и университетской библиотекой Орхуса)	NetarchiveSuite – это полный пакет программного обеспечения для веб-архивирования, разработанный в 2004 г. и используемый в производственной среде для сбора данных из датской сети с 2005 г. Основная функция NetarchiveSuite – планировать, составлять график и запускать сбор данных сети Интернет. Он масштабируется для решения широкого круга задач, от небольших	WARC	<a href="https://github.com/netarchivesuite">https://github.com/netarchivesuite</a>
	вместе с национальными библиотеками Франции, Австрии (ONB), Испании (BNE) и Швеции (KBSE)	тематических сборов (например, связанных со специальными событиями или специальными доменами) до сбора и архивирования контента всего национального домена. NetarchiveSuite построен на базе поискового робота Heritrix.		
HTTrack	Xavier Roche & other contributors	HTTrack – кроссплатформенный офлайн-браузер в открытом доступе. Позволяет загружать веб-сайты из Интернета на локальный компьютер [24].	HTML	<a href="http://www.httrack.com/">http://www.httrack.com/</a>
Rosetta	Proquest, ExLibris	Rosetta – это комплексная система для управления цифровыми активами и их хранения для библиотек, архивов, музеев и других учреждений [25].	–	<a href="https://exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/">https://exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/</a>

Название	Разработчики	Описание	Формат файла	Источник
WARCrefs	Библиотека Александрины	Инструмент дедупликации содержимого веб-архивов Библиотеки Александрины.	WARC	<a href="https://github.com/arcalex/warcrefs/blob/master/README.md">https://github.com/arcalex/warcrefs/blob/master/README.md</a>
IPFS	Protocol Labs при содействии open-source-сообщества	IPFS – это одноранговая распределённая файловая система, которая использует содержимое файла в качестве адреса для поиска.	–	<a href="https://ipfs.io/">https://ipfs.io/</a>
Umbra		Umbra – это инструмент автоматизации браузера, разработанный для службы веб-архивирования <a href="https://archive-it.org/">https://archive-it.org/</a> .	–	<a href="https://github.com/internetarchive/umbra">https://github.com/internetarchive/umbra</a>

### Отечественное веб-архивирование: состояние и перспективы развития

Особое внимание ресурсам, связанным с историей и развитием России, уделяет Президентская библиотека, работы по веб-архивированию в которой начались в 2017 г. При выборе ресурсов для веб-архивирования в первую очередь рассматривались приоритетные для библиотеки темы: история российской государственности и права и русский язык как государственный язык Российской Федерации. Поэтому для сохранения были выбраны сайты органов государственной власти. Сегодня на периодической основе Президентская библиотека архивирует такие ресурсы, как официальный сайт Президента России kremlin.ru и официальный сайт Правительства России government.ru. Стоит отметить, что новая архивная копия создаётся каждый день, это даёт возможность увидеть, как сайт менялся с течением времени [11].

С проектом сбора данных национального домена в единый электронный архив выступает автономная некоммерческая организация «Информационная культура» при поддержке международных стандар-

тов и использовании открытого программного обеспечения Heritrix в формате WARC. Размещённый на домене .org Национальный цифровой архив России на данный момент содержит БД «Консервированное государство», посвящённую аккаунтам в соцсетях, связанным с деятельностью органов государственной власти и её представителей.

В базе также содержатся все версии сайтов госучреждений РФ. Следующая часть данного проекта – «Специальные коллекции» – по версии разработчиков, должна будет содержать различные тематические коллекции, связанные с событиями, людьми, организациями и явлениями, не попадающими под критерии основных проектов цифрового архива. Помимо этого, согласно сайту архива, предлагается сбор данных по исторической, религиозной и другим областям знаний [12].

Специалистами РНБ Н. М. Балацкой и М. Б. Мартиросовой был исследован зарубежный опыт формирования национальных веб-архивов (в национальных библиотеках) и доступных краеведческих веб-архивов публичных и университетских библиотек. Были рассмотрены проблемы определения и формализации критериев отбора документов, создания методики отбора документов и описательных метаданных для файлов. При отборе ресурса для веб-архивирования был предложен индуктивный подход, при котором основной ценностью является документ, входящий в состав данного ресурса [13].

## **Выводы**

Наиболее актуальными проблемами, с которыми сталкиваются зарубежные библиотеки при формировании веб-архивов, являются интеграция процессов веб-архивирования в общую среду формирования фонда, вопросы авторского права при копировании материалов и технические аспекты фиксации сложных для копирования объектов. В качестве комплексного решения библиотеки либо обращаются к глобальным службам веб-архивирования (Internet Archive), либо сужают область формирования веб-архивов до небольших тематических коллекций.

Среди проектов зарубежных библиотек в этой области преобладает тенденция к сохранению «голоса эпохи»: стремление зафиксировать для фонда цифровое наследие в виде записей различных социальных групп и отдельных личностей, объединённых хронологически



ми или тематическими рамками. Данная тенденция нашла отражение в концепции веб-архивирования Британской национальной библиотеки, краеведческом проекте Community Webs, коллекциях национальных библиотек Ирландии и Австралии.

В России такого рода деятельность находится на начальной стадии разработки и сегодня в большей степени направлена на сохранение сегментов Интернета, связанных с государственной властью. Правительственные сайты являются общественным достоянием, поэтому копирование и архивация данного вида ресурсов, а также веб-страниц органов власти и отдельных её представителей могут производиться без заключения дополнительных соглашений.

Тем не менее отдельных нормативно-правовых актов, регламентирующих вопросы авторского права при веб-архивировании веб-сайтов, в современном отечественном законодательстве не обнаружено. Создание тематических коллекций и микроархивов возможно при соблюдении международных норм и стандартов цифрового копирования и предоставления доступа на условиях, согласованных с правообладателем. Согласно данным Консорциума по веб-архивированию, руководство крупнейших зарубежных библиотек, формирующих собственные веб-архивы, сталкивается с проблемой получения обратной связи от правообладателей при запросе разрешения на сбор, обработку и хранение данных, представленных в Сети.

Наиболее безопасной с правовой точки зрения является загрузка материалов, предоставляемых по лицензиям Creative Commons CC BY, CC BY-SA и CC0. Самой оптимальной для веб-архивирования является последняя лицензия, поскольку переводит произведение в режим, аналогичный общественному достоянию. В рамках краеведческой деятельности актуальным является опыт иностранных коллег по сохранению цифрового наследия жителей региона. Для реализации проектов в РНБ рекомендуется включение в профиль комплектования отдельной главы по методике отбора документов для селективного веб-архивирования с отражением критериев отбора документов.

## СПИСОК ИСТОЧНИКОВ

1. **Редькина Н. С.** Мировые тенденции развития веб-архивов библиотек // Научные и технические библиотеки. 2021. № 1. С. 99–114.
2. **ISO/TR 14873:2013(en)** Information and documentation – Statistics and quality issues for web archiving. URL: <https://www.iso.org/standard/55211.html> (дата обращения: 14.10.2020).
3. **International** Internet Preservation Consortium. URL: <https://netpreserve.org/> (дата обращения: 26.10.2020).
4. **List of Web** archiving initiatives // Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives#cite\\_ref-A\\_survey\\_on\\_web\\_archiving\\_initiatives\\_1-0](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#cite_ref-A_survey_on_web_archiving_initiatives_1-0) (дата обращения: 26.10.2020).
5. **Digital** Legal Deposit in Selected Jurisdictions / The Law Library of Congress, Global Legal Research Center. URL: <https://www.loc.gov/law/help/digital-legal-deposit/digital-legal-deposit.pdf> (дата обращения: 26.10.2020).
6. **Web Archiving** at National Libraries – Findings of Stakeholders. Consultation by the Internet Archive / Internet Archive, Global Web Services. URL: [https://archive.org/stream/InternetArchiveStakeholdersConsultationFindingsPublic/InternetArchiveStakeholdersConsultation-Findings\\_Public\\_djvu.txt](https://archive.org/stream/InternetArchiveStakeholdersConsultationFindingsPublic/InternetArchiveStakeholdersConsultation-Findings_Public_djvu.txt) (дата обращения: 02.11.2022).
7. **About** Community Webs. URL: <https://communitywebs.archive-it.org/about.html> (дата обращения: 26.01.2022).
8. **Lohndorf** Jillian Known Web Archiving Challenges / Jillian Lohndorf // Support.archive-it.org. URL: <https://support.archive-it.org/hc/en-us/articles/209637043-Known-Web-Archiving-Challenges> (дата обращения: 14.08.2021).
9. **Sawood** Alam Supporting Web Archiving via Web Packaging / Sawood Alam, Michele C. Weigle, Michael L. Nelson, Martin Klein, and Herbert Van de Sompel. URL: <https://arxiv.org/pdf/1906.07104.pdf> (дата обращения: 14.08.2021).
10. **The IETF** Datatracker. URL: <https://datatracker.ietf.org/wg/wpac/about/> (дата обращения: 26.06.2021).
11. **Президентская** библиотека первой в стране начала формирование сетевого архива. URL: <https://www.prlib.ru/news/689438> (дата обращения: 17.05.2021).
12. **Национальный** цифровой архив России. URL: <https://ruarxive.org/> (дата обращения: 26.10.2020).
13. **Балацкая Н. М., Мартиросова М. Б.** Веб-архивирование как задача национальной и краеведческой библиографии // Библиосфера. 2021. № 3. С. 12–17.
14. **Эйдемиллер И. В.** Комплектование фондов библиотек в условиях цифровой среды: основные тренды, вечные ценности и новые возможности: презентация. URL: <https://ru.calameo.com/read/00319731011e1086be577> (дата обращения: 11.07.2021).

15. **Браккер Н. В., Куйбышев Л. А.** Сбор и архивирование сетевых ресурсов. Опыт национальных библиотек зарубежных стран // Библиотекосведение. 2013. № 2. С. 88–96.

## References

1. **Red'kina N. S.** Mirov'e tendentsii razvitiia veb-arhivov bibliotek // Nauchny'e i tekhnicheskie biblioteki. 2021. № 1. S. 99–114.
2. **ISO/TR 14873:2013(en)** Information and documentation – Statistics and quality issues for web archiving. URL: <https://www.iso.org/standard/55211.html> (data obrashcheniia: 14.10.2020).
3. **International** Internet Preservation Consortium. URL: <https://netpreserve.org/> (data obrashcheniia: 26.10.2020).
4. **List of Web** archiving initiatives // Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives#cite\\_ref-A\\_survey\\_on\\_web\\_archiving\\_initiatives\\_1-0](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#cite_ref-A_survey_on_web_archiving_initiatives_1-0) (data obrashcheniia: 26.10.2020).
5. **Digital** Legal Deposit in Selected Jurisdictions / The Law Library of Congress, Global Legal Research Center. URL: <https://www.loc.gov/law/help/digital-legal-deposit/digital-legal-deposit.pdf> (data obrashcheniia: 26.10.2020).
6. **Web Archiving** at National Libraries – Findings of Stakeholders. Consultation by the Internet Archive / Internet Archive, Global Web Services. URL: [https://archive.org/stream/InternetArchiveStakeholdersConsultationFindingsPublic/InternetArchiveStakeholdersConsultation-Findings\\_Public\\_djvu.txt](https://archive.org/stream/InternetArchiveStakeholdersConsultationFindingsPublic/InternetArchiveStakeholdersConsultation-Findings_Public_djvu.txt) (data obrashcheniia: 02.11.2022).
7. **About** Community Webs. URL: <https://communitywebs.archive-it.org/about.html> (data obrashcheniia: 26.01.2022).
8. **Lohndorf** Jillian Known Web Archiving Challenges / Jillian Lohndorf // Support.archive-it.org. URL: <https://support.archive-it.org/hc/en-us/articles/209637043-Known-Web-Archiving-Challenges> (data obrashcheniia: 14.08.2021).
9. **Sawood** Alam Supporting Web Archiving via Web Packaging / Sawood Alam, Michele C. Weigle, Michael L. Nelson, Martin Klein, and Herbert Van de Sompel. URL: <https://arxiv.org/pdf/1906.07104.pdf> (data obrashcheniia: 14.08.2021).
10. **The IETF** Datatracker. URL: <https://datatracker.ietf.org/wg/wpack/about/> (data obrashcheniia: 26.06.2021).
11. **Prezidentskaia** biblioteka pervoi` v strane nachala formirovanie setevogo arhiva. URL: <https://www.prlib.ru/news/689438> (data obrashcheniia: 17.05.2021).
12. **Natsional'ny'i`** tcifrovoy` arhiv Rossii. URL: <https://ruarxive.org/> (data obrashcheniia: 26.10.2020).
13. **Balatckaia N. M., Martirosova M. B.** Veb-arhivirovanie kak zadacha natsional'noi` i kraevedcheskoi` bibliografii // Bibliosfera. 2021. № 3. S. 12–17.

14. **E`i`demiller I. V.** Komplektovanie fondov bibliotek v usloviiahk tcifrovoy` srede`: osnovny`e trendy`, vechny`e cennosti i novy`e vozmozhnosti: prezentatciia. URL: <https://ru.calameo.com/read/00319731011e1086be577> (data obrashcheniia: 11.07.2021).
15. **Brakker N. V., Kui`by`shev L. A.** Sbor i arhivirovanie setevy`kh resursov. Opy`t natsional`ny`kh bibliotek zarubezhny`kh stran // Bibliotekovedenie. 2013. № 2. S. 88–96.

### **Информация об авторе / Information about the author**

**Смирнов Артём Александрович** – методист I категории, сектор изучения библиотечных фондов научно-методического отдела Российской национальной библиотеки, Санкт-Петербург, Российская Федерация  
artjomvajda@gmail.com

**Artyom A. Smirnov** – First Category Specialist, Sector for Library Collection Studies, Research and Instructional Department, National Library of Russia, St. Petersburg, Russian Federation  
artjomvajda@gmail.com