

УДК 015:004

<https://doi.org/10.20913/1815-3186-2021-3-12-17>

Веб-архивирование как задача национальной и краеведческой библиографии¹

Н. М. Балацкая[✉], М. Б. Мартиросова

**Балацкая
Надежда Михайловна,**

Российская
национальная
библиотека,
ул. Садовая, 18, Санкт-
Петербург, 191069,
Россия,
кандидат

педагогических наук, старший
научный сотрудник
отдела библиографии и краеведения

ORCID: [0000-0001-5770-2757](https://orcid.org/0000-0001-5770-2757)e-mail: kray@nlr.ru

**Мартиросова
Маргарита Бедросовна,**

Донская
государственная
публичная библиотека,
ул. Пушкинская,
175А, Ростов-на-Дону,
344006, Россия,
заведующий отделом

краеведения

ORCID: [0000-0002-5819-2338](https://orcid.org/0000-0002-5819-2338)e-mail: donkray@yandex.ru

Аннотация. В связи с разработкой модели и методики электронных краеведческих архивов для российских библиотек проанализирован зарубежный опыт формирования национальных веб-архивов (в национальных библиотеках) и доступных краеведческих веб-архивов публичных и университетских библиотек. Показано, что глобальная деятельность по сохранению сетевого контента разделена на национальные и региональные сегменты и является национальной и краеведческой по содержанию. Цели и вся структура ее полностью укладываются в классические формы и направления национальной/краеведческой библиографии, играющей ключевую роль в процессах отбора и библиографической обработки нового класса источников. Для отбора применяются все характерные для национальной/краеведческой библиографии критерии связи со страной/регионом/территорией: место производства ресурса, его национальный язык и содержание; сам отбор осуществляется специалистами библиотек. Выявлены причины, по которым подготовка архивируемого сетевого контента к использованию, независимо от возможностей автоматизированных технологий, требует привлечения специалистов-библиографов, применения классических инструментов и методов библиографического описания и индексирования содержания. Рассмотрен вопрос о выборе объектов для архивирования (сайтов, подкаталогов, отдельных документов). Показаны особое значение, преимущества и проблемы работы с отдельными документами. Обоснована необходимость построения таких стратегий веб-архивирования, в которых «отправными точками» являются отдельные документы, а сайты оцениваются и отбираются по продуктивности в качестве их «контейнеров». С учетом мировых тенденций и условий российских библиотек предложена модель веб-архивирования, внедрение которой, в отсутствие национального репозитория и современных инструментов веб-архивирования, позволит начать накопление необходимого опыта и перестать терять ценный сетевой контент.

Ключевые слова: веб-архивирование, национальная библиография, краеведческая библиография, библиотеки, сетевые источники, отбор документов, библиографическая обработка

Для цитирования: Балацкая Н. М., Мартиросова М. Б. Веб-архивирование как задача национальной и краеведческой библиографии // Библиосфера. 2021. № 3. С. 12–17. <https://doi.org/10.20913/1815-3186-2021-3-12-17>.

Статья поступила в редакцию 01.07.2021

Получена после доработки 27.08.2021

Принята для публикации 30.08.2021

© Н. М. Балацкая, М. Б. Мартиросова, 2021

¹ Научная статья подготовлена на основании доклада, сделанного на III Международном библиографическом конгрессе, 27–30 апреля 2021 г.

Web Archiving as a Task of National and Local History Bibliography

Nadezhda M. Balatskaya , Margarita B. Martirosova

Balatskaya Nadezhda Mihaylovna,
National Library of Russia,
Sadovaya str., 18, Saint-Petersburg,
191069, Russia,
PhD, Senior Researcher of the
Department of Bibliography
and Local History
ORCID: [0000-0001-5770-2757](https://orcid.org/0000-0001-5770-2757)
e-mail: kray@nlr.ru

Martirosova Margarita Bedrosovna,
Don State Public Library,
Pushkinskaya str., 175A, Rostov-on-Don,
344006, Russia,
Head of the Department of Local Studies
ORCID: [0000-0002-5819-2338](https://orcid.org/0000-0002-5819-2338)
e-mail: donkray@yandex.ru

Received 01.07.2021
Revised 27.08.2021
Accepted 30.08.2021

Abstract. The foreign experience of forming the majority of national web archives (in national libraries) and many accessible local history archives of public and university libraries has been analyzed due to the development of a model and methodology for electronic local history archives for Russian libraries.

It is shown that the global activity on preserving network content is divided into national and regional segments and is national and local history in both contents. Its goals and structure fully fit into the classical forms and directions of national/local history bibliography, which turns out to be the key in all processes of selection and bibliographic processing of a new class of sources. For selection, all the criteria of connection with the country/region/territory, the place of production of the resource, its national language and content, characteristic of the national/local history bibliography; the selection itself is carried out by library specialists. The analysis has revealed the reasons why the preparation of the archived online content for use, regardless of the capabilities of automated technologies, in practice also requires the involvement of specialist in bibliography, the use of classical tools and methods of bibliographic description and indexing of content. The issue of choosing objects for archiving (sites, subdirectories, individual documents) is specially considered. The special significance, advantages and problems of treating individual documents are shown. The necessity of forming such web archiving strategies is substantiated, in which “starting points” are separate documents, and sites are evaluated and selected for productivity as their “containers”. Taking into account the global trends and the conditions of Russian libraries, a model of web archiving is proposed, the implementation of which, in the absence of a national repository and modern tools, will allow the beginning of accumulating some necessary experience and valuable network content.

Keywords: web archiving, national bibliography, local history bibliography, libraries, web sources, document selection, bibliographic processing

Citation: Balatskaya N. M., Martirosova M. B. Web Archiving as a Task of National and Local History Bibliography. *Bibliosphere*. 2021. № 3. P. 12–17. <https://doi.org/10.20913/1815-3186-2021-3-12-17>.

Введение

Наш интерес к теме веб-архивирования связан с разработкой модели и методики создания электронных краеведческих архивов в российских региональных библиотеках (*Руководство...*, 2017, 2018). В статье представлены некоторые результаты анализа зарубежного опыта веб-архивирования в национальных, публичных и университетских библиотеках.

Громадное количество документов (в том числе ценных и уникальных) рождается, живет и исчезает в Сети (*Jones et al., 2016*). Расхожее мнение о том, что интернет помнит всё, опровергается данными специалистов: средняя продолжительность жизни веб-страницы до ее изменения или удаления составляет всего 100 дней (*Graham, 2016*); 80 % всех веб-сайтов больше не будут доступны в той же форме или полно-

стью исчезнут в течение года². Поэтому уже четверть века и в отдельных странах, и на международном уровне³ разрабатываются стратегии и инструменты веб-архивирования.

Роль библиотек и библиографии в веб-архивировании

Глобальная проблема сохранения Сети решается сейчас путем разделения на национальные сегменты – как сумма задач сохранения национального культурного и исторического наследия. Создание таких веб-архивов почти везде поручено национальным библиотекам, поскольку именно они обладают необходимыми для этого опытом и инструментами (*Bracciotti, 2019*).

² Luxembourg Web Archive / Bibliothèque nationale du Luxembourg. URL: <https://www.webarchive.lu> (accessed 14.02.2021).

³ List of Web archiving initiatives. URL: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives (accessed 11.06.2021).

На другом уровне – региональном – в последнее время появляются краеведческие веб-архивы. Их создают публичные библиотеки, видящие в этом необходимое развитие своей краеведческой деятельности в цифровой среде⁴, а также некоторые университетские – для обеспечения региональных исследований своих ученых.

Таким образом, задача веб-архивирования является национальной и краеведческой как по содержанию, так и по самой своей постановке. При этом цели и вся структура этой работы полностью укладываются в формы и направления библиографии – и можно сказать точнее – в формы и направления краеведческой⁵ и неотличимой от нее на негосударственном уровне национальной библиографии⁶.

В масштабной деятельности по веб-архивированию, объединяющей и программно-технические, и правовые аспекты⁷, библиографическая составляющая не просто присутствует, а отвечает за смыслы и содержание и является ключом к работе по отбору, библиографической обработке и продвижению нового класса источников.

Библиографический подход в отборе материалов для веб-архивирования

Начнем с отбора: исчерпывающая полнота все чаще осознается как не только недостижимая, но и неверная цель (Storti, 2019).

В общей постановке целей веб-архивирования как сохранения и обеспечения доступности национального сегмента или краеведческого контента Сети выделяются три классических направления национальной и краеведческой библиографии:

1. Выявление ресурсов, связанных со страной или регионом местом публикации, производства или распространения.

2. Выявление ресурсов на национальном языке или языках.

3. Выявление ресурсов, связанных со страной или регионом содержанием (независимо от места публикации, производства и распространения и от языка).

При всей обеспеченности современными технологиями даже первая, более простая задача не решается автоматически, поскольку и наилучшим образом организованный сбор сайтов

⁴ Empowering cultural heritage organizations to create community history web archives. *Community webs*. URL: <https://communitywebs.archive-it.org/index.html> (accessed 11.06.2021).

⁵ About us. *PADICAT the web archive of Catalonia*. URL: <https://www.padicat.cat/en/about-us> (accessed 11.06.2021).

⁶ Info // Finnish Web Archive. URL: <https://verkkoarkisto.kansalliskirjasto.fi/va/info> (accessed 11.06.2021).

⁷ Legal issues / International Internet Preservation Consortium – IIPC. URL: <https://netpreserve.org/web-archiving/legal-issues> (accessed 10.06.2021).

по доменным именам не обеспечивает полноты сбора: через это «сито» свободно проваливаются сайты, принадлежащие частным лицам и любым сообществам. Вероятно, поэтому даже при наличии такого сбора ни одна из национальных библиотек, формирующих свои веб-архивы, не ограничивается им, а использует и другие критерии⁸. И практически все они, кроме национальных доменных имен и, может быть, национального языка, требуют дополнительного «человеческого» выбора, опирающегося на внешние источники – например, принадлежность сайтов каким-то категориям государственных, научных, культурных, образовательных учреждений, гражданство или место жительства владельцев, предназначенность для населения страны или региона⁹.

А для третьей задачи, связанной с отбором веб-ресурсов по содержанию (то есть о стране или регионе), таких трудно формализуемых критериев значительно больше. Например:

- значимость для сохранения национальной культуры;
- соответствие интересам национальной/местной аудитории;
- мера связанности с местной жизнью;
- полнота представленности разных групп местного населения;
- артефактная, институциональная и научная ценность контента;
- уникальность контента;
- актуальность контента;
- востребованность контента;
- известность создателя сайта;
- и, наконец, собственно качество – сайта и самого контента.

Применение этих критериев еще сложнее, с еще большей необходимостью опоры на краеведческие источники, знания и профессионализм специалистов именно в области национальной и краеведческой библиографии.

Библиографическая обработка архивируемых материалов

Библиографическая обработка сетевых материалов не проще, чем их отбор. В зарубежных публикациях по проблемам веб-архивирования подчеркивается, что его целью является не просто сбор «сырого» контента, а сохранение в подготовленном к использованию виде. Есть причины, по которым не удастся сделать это автоматически.

⁸ How it works / Bibliothèque nationale du Luxembourg. URL: <https://www.webarchive.lu/how-it-works> (accessed 25.06.2021).

⁹ Web site harvesting / Bibliothèque et Archives nationales du Québec. URL: https://www.banq.qc.ca/services/bibliotheque_nationale/depot_legal/collecte_sites/index.html?language_id=1 (accessed 25.06.2021).

Прежде всего, это трудности с описательными метаданными (то есть формальным описанием), отмечаемые многими специалистами¹⁰. При веб-архивировании снимаются самые сложные, практически неразрешимые проблемы, связанные с изменчивостью и эфемерностью сетевого контента как объекта описания. «Остановленное мгновение» его жизни можно описать классическим образом, а сам «живой» сайт, по всей видимости, и не должен рассматриваться в качестве объекта библиографического описания. Но остаются трудности, обусловленные неполнотой или искажением сведений, необходимых для атрибуции сетевых источников. Они пока не имеют стандартных практических решений и требуют индивидуального подхода профессионального библиографа. Об этом уже много сказано; из недавних публикаций сошлемся на рекомендации рабочей группы по метаданным для веб-архивирования OCLC (Dooley, Bowers, 2018).

И так же не удастся полностью автоматизировать работу по индексированию содержания архивируемых материалов. В наиболее общем виде ситуация хорошо известна, представлена во многих публикациях по проблемам веб-архивирования и заключается в очевидной недостаточности (при всей ценности!) возможностей автоматизированного контекстного поиска, равно как и применения «народной библиографии», когда индексы содержания в той или иной форме создает сам автор (Smith-Yoshimura, 2020). На практике:

- таких индексов часто вообще нет (они не заложены авторами);
- они могут быть неполными, неправильными (некорректно сделаны авторами);
- семантическая микроразметка развивается, но не является обязательной;
- метаданные создаются авторами сетевых ресурсов с целями, часто не совпадающими с целями потенциальных пользователей;
- необходимая информация может оказаться не главной в размечаемом документе (типичная картина для краеведения), и тогда она с высокой степенью вероятности не будет найдена и сохранена.

Поэтому практически во всех известных нам веб-архивах при существующих возможностях автоматизированного контекстного поиска ведется и традиционная библиографическая обработка архивируемых материалов, включающая их библиографическое описание и раскрытие содержания – с применением рубрикаторов, ключевых слов и предметных рубрик. Более того, создатели многих национальных веб-архивов

специально подчеркивают ценность «ручной» (то есть тщательной и качественной) обработки (Storti, 2019). Есть серьезные причины не считать это временным положением или инерцией традиционной библиотечной практики, не успевающей за развитием цифровых технологий.

Добавим, что специфика национальной и краеведческой библиографии в содержательной обработке веб-контента не очевидна для тех, кто ею не занимается, поскольку все используемые инструменты индексирования принципиально стандартны. Здесь все дело в особом взгляде на материал. Коль скоро сбор «своего» сетевого контента, по крайней мере по основной группе критериев, осуществляется именно по содержанию – национальному или краеведческому («о Чехии», «о Тюрингии», «о Каталонии»), тот же подход со всей очевидностью переходит и на методику содержательной библиографической обработки: связь сетевого ресурса со страной или регионом нужно не только увидеть при отборе, но и показать для потенциального пользователя.

Элементы контента как объект архивирования

В рамках рассматриваемой темы особого внимания заслуживает вопрос о выборе объектов для архивирования. Ими могут быть и целые сайты, и элементы контента: отдельные публикации, подборки, подкаталоги. Об этом втором уровне, продолжающем классическую библиографическую аналитику, стоит сказать подробнее. Возможность работы с отдельными публикациями, подборками и подкаталогами уже на этапе отбора контента специально предусмотрена во всех известных нам методиках (см., например, Rollason-Cass, 2021), и такие элементы контента в разных объемах присутствуют в большинстве национальных и особенно краеведческих веб-архивов.

В связи с высокой рассеянностью материала и высоким уровнем точности пользовательских задач работа с отдельными документами традиционно доминирует в национальной и краеведческой библиографии – в той ее важнейшей части, которая связана со страной или регионом содержанием. Применительно к задачам сохранения Сети аналитика имеет ряд преимуществ, связанных с компактностью хранения, принципиальной возможностью точной атрибуции и качественного описания, легкостью интеграции в структуру краеведческих ресурсов и «экспертной» ценностью для пользователей. В сущности, именно конкретный документ (а не целый сайт или блог) и нужен пользователю.

Особенности работы с сетевой «аналитикой» почти не рассматриваются в литературе

¹⁰ Web Archiving Metadata Working Group. URL: <https://www.oclc.org/research/areas/research-collections/wam.html> (accessed 25.06.2021).

и дискуссиях по проблемам веб-архивирования. Между тем именно здесь сконцентрированы все вышеупомянутые трудности библиографического описания сетевых материалов (Балацкая, Мартиросова, 2019). И здесь мы сталкиваемся с новым типом отношений между источником (сайтом) и элементами его контента, не укладывающимся в модель классического аналитического описания, и с новой номенклатурой самих этих элементов, которые могут становиться самостоятельными объектами в составе веб-архива. Разработка методики аналитического библиографического описания отдельных сетевых документов в упомянутом руководстве WAM OCLC была отнесена к числу предстоящих исследовательских задач группы (Dooley, Bowers, 2018, p. 32–33).

Аналитика присутствует во многих архивных коллекциях наряду с сайтами, но нигде не представляет самостоятельный слой источников. Мы же видим здесь необходимость построения таких стратегий веб-архивирования, в которых «отправными точками» являются отдельные документы, а сайты оцениваются и отбираются по продуктивности в качестве их «контейнеров» (Storti, 2019).

Выводы

Таким образом, при всей новизне задачи веб-архивирования и освоения совершенно нового класса информационных источников здесь решающее значение имеет опыт классической краеведческой и национальной библиографии, накопленный в библиотеках. Остается добавить еще один аргумент, который все чаще звучит в дискуссиях по этим проблемам. Он заключается в том, что недостаточно отобрать и обработать материал для веб-архива – нужно еще интегрировать его в уже существующий корпус печатных, архивных документов¹¹. Объединяющим инструментом может быть (и на практике является) именно библиография – через единство подхода к формированию целостного корпуса источников и через метаданные, которые могли бы обеспечивать единый «вход» в него¹².

Мы понимаем, что задача создания единого национального веб-архива в такой большой стране, как Россия, невыполнима силами любого федерального центра, независимо от его статуса и мощности – по той простой причине, что специалисты в ключевых вопросах отбора и обработки сетевого контента, владеющие знаниями источников, находятся на другом уровне – региональном. Поэтому нам кажется правильным начать эту работу с регионального уровня, не дожидаясь решений на государственном. Здесь самое главное – строить эту модель так, чтобы позднее она без ломки и перестройки вросла в целостную общегосударственную, если/когда она будет у нас построена.

С учетом выявленных тенденций и реальных условий российских библиотек нами предложена модель веб-архивирования (Балацкая, Мартиросова, 2020), ключевыми особенностями которой являются:

- распределение этой задачи между региональными библиотеками, с возможностью сквозного поиска по метаданным;
- экспертный подход к отбору материалов на основе системы критериев, с одной стороны, общих для сетевых и традиционных краеведческих документов, с другой – детально разработанных применительно к различным видам сетевых документов и (на другом уровне) к их источникам;
- приоритет работы с элементами контента (документами) независимо от возможностей архивирования целых сайтов;
- применение классических инструментов раскрытия содержания независимо от возможностей контекстного поиска;
- интеграция архивируемых сетевых ресурсов с печатными и архивными краеведческими источниками во всех аспектах сбора, хранения, поиска и продвижения.

Внедрение этой модели уже теперь, в отсутствие национального репозитория и современных инструментов веб-архивирования, позволит начать накопление необходимого опыта и перестать терять ценный контент.

¹¹ Overview / National Széchényi Library // Rákóczi webarchivum. URL: <https://rakoczi2019.webarchivum.oszk.hu/en/introduction> (accessed 25.06.2021).

¹² National Library of New Zealand catalog / The National Library of New Zealand. URL: <https://natlib-primo.hosted.exlibrisgroup.com/primo-explore/search?vid=NLNZ&sortby=rank> (accessed 25.06.2021).

Список источников / References

- Балацкая Н. М., Мартиросова М. Б. О разработке новых инструментов краеведческой работы в цифровой среде // Вестник Национальной библиотеки Республики Саха (Якутия). 2020. № 2. С. 40–46 [Balatskaya NM and Martirosova MB (2020) On developing new tools for local history work in the digital environment. *Vestnik Natsional'noi biblioteki Respubliki Sakha (Yakutiya)* 2: 40–46. (In Russ.)].
- Балацкая Н. М., Мартиросова М. Б. Сетевые ресурсы удаленного доступа как объект библиографического описания: к выходу в свет ГОСТ 7.1.100–2018 // Библиография. 2019. № 2. С. 16–26 [Balatskaya NM and Martirosova MB (2019) Remote access network resources as an object of bibliographic description: to edition of GOST 7.1.100-2018. *Bibliografiya* 2: 16–26. (In Russ.)].
- Руководство по краеведческой деятельности общедоступных (публичных) библиотек РФ : принято на Всерос. библиограф. конгр. (XXIII Ежегод. конф. РБА), Владимир, 18 мая 2018 г. / сост.: Н. М. Балацкая, Е. Н. Ефремова, А. М. Ковалева, М. Б. Мартиросова // Библиотека и закон. 2018. Вып. 44. С. 219–238 [Balatskaya NM, Efremova AM, Kovaleva AM and Martirosova MB (comps) (2018) Guide to the local history activities of public libraries of the Russian Federation: accepted for All-Russ. bibl. Congr. (XXIII Annu. RBA conf.), Vladimir, May 18, 2018. *Biblioteka i zakon* 44: 219–238. (In Russ.)].
- Руководство по краеведческой деятельности центральной библиотеки субъекта РФ : принято на Всерос. библиограф. конгр., Красноярск, 18 мая 2017 г. / сост.: Н. М. Балацкая, А. М. Ковалева, М. Б. Мартиросова // Библиотека и закон. 2017. Вып. 42. С. 57–71 [Balatskaya NM, Kovaleva AM and Martirosova MB (comps) (2017) Guide to local history activities of the central library of the Russian Federation subject: accepted for All-Russ. bibl. Congr., Krasnoyarsk, May 18, 2017. *Biblioteka i zakon* 42: 57–71. (In Russ.)].
- Bracciotti L (2019). Il web archiving. Conservazione e uso di una nuova fonte. *Officina Della Storia* Gen. 10. URL: <https://www.officinadellastoria.eu/it/2019/01/10/il-web-archiving-conservazione-e-uso-di-una-nuova-fonte> (accessed 11.06.2021).
- Dooley J and Bowers K (2018) Descriptive metadata for web archiving: recommendations of the OCLC research library partnership web archiving metadata working group. Dublin: OCLC Research. URL: <https://doi.org/10.25333/C3005C> (accessed 11.06.2021).
- Graham M (2016) Why the Internet needs the wayback machine, the site that archives the web. *Entrepreneur* Aug. 4. URL: <https://www.entrepreneur.com/article/280107> (accessed 12.01.2021).
- Jones SM, Van de Sompel H, Shankar H, Klein M, Tobin R and Grover C (2016) Scholarly context adrift: three out of four URI references lead to changed content. *PLoS One* Dec. 2. URL: <https://doi.org/10.1371/journal.pone.0167475> (accessed 25.06.2021).
- Rollason-Cass S (2021) How seeds, documents, and collections work together. *Archive-It User Guide*. URL: <https://support.archive-it.org/hc/en-us/articles/360000706843-How-seeds-documents-and-collections-work-together> (accessed 25.06.2021).
- Smith-Yoshimura K (2020) Transitioning to the next generation of metadata : OCLC research report. Dublin: OCLC Research. URL: <https://doi.org/10.25333/rqgd-b343> (accessed 25.06.2021).
- Storti C (2019) Web archiving, “sfida culturale”: il servizio della Biblioteca Nazionale Centrale di Firenze. *FPA Digital360* Giugno 12. URL: <https://www.forumpa.it/pa-digitale/gestione-documentale/web-archiving-sfida-culturale-il-servizio-della-biblioteca-nazionale-centrale-di-firenze> (accessed 11.06.2021).