

Государственное бюджетное учреждение культуры Архангельской области
«Архангельская областная научная ордена «Знак Почета»
библиотека имени Н. А. Добролюбова»

Отдел краеведения «Русский Север»

СОЗДАНИЕ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ (КОЛЛЕКЦИИ)

ВЫПУСК 2

ТЕХНОЛОГИЯ ОЦИФРОВКИ БИБЛИОТЕЧНЫХ ФОНДОВ

Методические рекомендации

Архангельск
2017

Создание электронной библиотеки (коллекции). Технология оцифровки библиотечных фондов. Вып.2. : методические рекомендации / сост. Т. В. Фролова ; ГБУК АО Архангельская областная научная ордена «Знак Почета» библиотека им. Н. А. Добролюбова ; Отдел краеведения «Русский Север» . – Архангельск . – 2017 . – 13 с.

Оглавление

От составителей.....	4
Термины и определения.....	4
Подготовительная работа с документом, предназначенным для оцифровки.....	5
Сканирование документа.....	5
Обработка электронных документов. Создание архивных и пользовательских копий	7
Запись электронных копий на носители длительного хранения	8
Приложение № 1.....	9
Приложение № 2.....	10
Приложение № 3.....	11
Приложение № 4.....	12
Приложение № 5.....	13

От составителей

Данные методические рекомендации продолжают раскрывать тему «Создание электронных библиотек (коллекций)», актуальную для развития общедоступных библиотек. В первом выпуске методических рекомендаций¹ были рассмотрены вопросы, касающиеся отбора документов из библиотечных фондов для оцифровки, выбора оборудования и программного обеспечения для сканирования и обработки документов. В данном выпуске рассматриваются технологии сканирования документов, создания и обработки архивных и пользовательских копий.

Методические рекомендации основаны на опыте Архангельской областной научной библиотеки имени Н. А. Добролюбова (далее — АОНБ) по формированию Электронной краеведческой библиотеки «Русский Север» (далее — ЭКБ). Основой для их разработки стали локальные нормативные документы АОНБ (Концепция ЭКБ «Русский Север», «Положение о формировании и развитии ЭКБ «Русский Север»), а также «Рекомендации по оцифровке материалов из фондов библиотек», подготовленные Российской национальной библиотекой².

Основные термины и определения, связанные с данного рода деятельностью библиотек и использованные в методических рекомендациях, приведены на основании ГОСТа Р7.0.96-2016 «Электронные библиотеки. Основные виды. Структура. Технология формирования», введенного в действие с 01.07.2017 г.

Термины и определения

Архивная копия (мастер-копия) — цифровая копия, выполненная с максимальным качеством и с минимальной последующей обработкой.

Коллекция электронных документов — часть фонда электронной библиотеки, представляющая логическое объединение электронных документов, отобранных по какому-либо критерию принадлежности, и предназначенная для реализации задач, которые определены политикой электронной библиотеки.

Оператор электронной библиотеки — гражданин или юридическое лицо, осуществляющее деятельность по эксплуатации информационной системы, в том числе по обработке информации.

Пользовательская (экранная) копия — цифровая копия, предназначенная для просмотра на персональном компьютере локальными и удаленными пользователями, в формате PDF.

Фонд электронной библиотеки — упорядоченное собрание объектов, формируемое оператором электронной библиотеки и предназначенное для долговременного хранения и использования.

Электронная библиотека — информационная система, предназначенная для организации и хранения упорядоченного фонда электронных объектов и обеспечения доступа к ним с помощью единых средств навигации и поиска.

¹ Создание электронной библиотеки (коллекции). Оцифровка фондов муниципальных библиотек. С чего начать? Вып. 1. [Электронный ресурс] : методические рекомендации / сост. Т. В. Фролова // Библиотеки Архангельской области. – URL : <http://biblioteka29.ru/upload/medialibrary/2fa/2fa60c410ad862c0557627fd4b103e6d.doc> (Дата обращения: 22.12.2017).

² Рекомендации по оцифровке материалов из фондов библиотек : утв. 23.10.2013 г. Экспертным советом по цифровым копиям Российской национальной библиотеки. – URL : http://www.nlr.ru/pro/inv/digit_copy/recom.pdf (Дата обращения: 25.12.2017).

Электронный документ — документ в цифровой форме, для использования которого необходимы средства вычислительной техники или иные специализированные устройства для воспроизведения текста, звука, изображения.

Подготовительная работа с документом, предназначенным для оцифровки

Предварительная работа с документом, запланированным для оцифровки, включает в себя несколько этапов. При наличии дублетных экземпляров нужно выбрать наиболее подходящий для оцифровки (имеющий оригинальную обложку, автографы, пометы и т. п.). Затем необходимо просмотреть каждый лист выбранного документа на предмет полноты наличия или повреждения в нем страниц, иллюстраций, карт, таблиц и пр., и, если требуется, осуществить поиск недостающих частей документа в других книгохранилищах, в т. ч. сделать заказ по межбиблиотечному абонементу.

Для удобства дальнейшей работы с документом ему присваивается уникальный идентификационный номер, который вносится в форму «Список документов для оцифровки...» (Приложение № 1) в столбец «Примечания». Например, в АОНБ им. Н. А. Добролюбова идентификатор формируется из номера года, буквенного кода вида документа и порядкового номера документа за текущий год³. Например, 17kp010, где 17 — это 2017 год, kp — книга, 010 — 10-й документ, отсканированный в 2017 году.

В крупных библиотеках, где деятельность по оцифровке фонда осуществляется несколькими сотрудниками, для удобства организации работы на каждый документ заводится «Бланк сопровождения документа»⁴, в котором ставятся отметки о прохождении этим документом этапов оцифровки и об особенностях его электронных копий.

Нумерацию файлов необходимо начинать с обложки документа (с первой страницы первого номера, если это газета). Например, первый файл будет изображением обложки книги и получит имя 17bk001_0001.tif, где: 17bk001 — уникальный идентификатор объекта оцифровки, _0001 — порядковый номер файла (электронной копии страницы документа), tif — расширение файла (графический формат TIFF [tagged image file format]). Номер электронной копии формируется исходя из общего предполагаемого количества файлов по определённой схеме, представленной в таблице:

Количество страниц в документе меньше или равно	Нумерация начинается с
9	17bk001_1.tif
99	17bk001_01.tif
999	17bk001_001.tif
9999	17bk001_0001.tif

Сканирование документа

Для обеспечения качества электронной копии необходимо правильно задать параметры и режимы сканирования.

Прежде всего, необходимо определить требуемое **разрешение изображения**, измеряющееся в dpi (dots per inch — количество точек-пикселей на дюйм). Использовать разрешение ниже чем 300 dpi нецелесообразно. В случае если текст сканируемых страниц набран мелким шрифтом, то имеет смысл проводить сканирование с более высоким разрешением (400 или 600 dpi).

³ Правила формирования идентификационного номера электронного документа представлены в Приложении № 2

⁴ Пример бланка представлен в Приложении № 3

В таблице приведены рекомендуемые границы разрешения, применяемые при оцифровке объектов различных видов:

Вид объекта	Разрешение
Отдельные фотографии	максимально возможное в диапазоне 300–600 dpi
Фотоальбом	страница целиком — порядка 300 dpi, каждая фотография по отдельности — максимально возможное в диапазоне 300–700 dpi
Текст без рисунков	300–450 dpi
Текст с рисунками	страница целиком — 300–450 dpi, каждый рисунок по отдельности — 300–600 dpi
Отдельные гравюры	300–1200 dpi
Альбом с гравюрами	страница целиком — 300–450 dpi, каждая гравюра по отдельности — 300–600 dpi
Карты	300–600 dpi

Все элементы одного документа должны быть отсканированы с одинаковым разрешением.

Нужно помнить: чем выше разрешение изображения, тем больше в конечном счете будет размер файла (электронной копии страницы), и, как результат, увеличится объем всего электронного документа.

Не менее существенное значение имеет и правильный выбор **режима сканирования**. Три основные из них — цветной, оттенки серого и черно-белый.

Цветной режим (color) — это получение картинка с глубиной цвета в 24 бит, т. е. изображения со всеми возможными переходами оттенков. Данный режим сканирования наиболее подходит для страниц, содержащих цветные иллюстрации и разноцветные надписи. Он может понадобиться и в том случае, если необходимо сохранить не только сам текст, но и внешний вид его носителя — фактуру и цвет бумаги, какие-нибудь дефекты типа царапин, чьих-то рукописных пометок, обтрепавшихся от времени уголков страниц и т. п.

Оцифровка в **режиме оттенков серого (grayscale)** — это получение электронной копии с глубиной цвета 8 бит. Визуально такой документ выглядит как черно-белая фотография. Этот режим сканирования хорош для страниц с черно-белыми иллюстрациями, содержащих различные оттенки серого цвета, чаще всего это газеты.

Третий режим сканирования называется **черно-белым (black and white)**, еще он именуется одноцветным или монохромным. Его не следует путать с режимом оттенков серого, в черно-белом режиме получаемый электронный документ имеет глубину цвета 1 бит. При сканировании светлые участки страницы воспринимаются как имеющие только белый цвет, а темные — только черный. Такой режим наиболее подходит для страниц, содержащих текст, таблицы и рисунки, представляющие собой совокупности контрастных, четко очерченных линий, например графики и схемы.

Выбор режимов и разрешения сканирования во многом зависит от исходного документа, предназначенного для оцифровки, особенностей проекта, целей и задач, которые ставит перед собой библиотека. Чаще всего для оцифровки библиотечных фондов, в том числе газет, используется режим grayscale. Если в оцифровываемом документе присутствует значимая цветная информация (более 25 %), то сканирование необходимо производить в режиме color.

Для того чтобы сформировать качественную электронную коллекцию, библиотеке необходимо обеспечивать высокое качество электронных копий. Документы должны быть

отсканированы полностью, включая лицевую и тыльную сторону обложки, титульный лист и оборот титульного листа, концевой титульный лист (колофон), оборотную сторону переплётной крышки, вклейки и оборотные стороны вклеек и т. д.

Пустые страницы (вакаты) также подлежат сканированию, независимо от наличия или отсутствия на них номера и другой значащей информации.

Развороты вшитых или вклеенных в оригинал листов, на которых расположена логически цельная информация (распашная карта, изображение, несколько изображений под одним названием, таблица с переходящими строками и т. д.), должны быть представлены как единое изображение (разворотом).

Цифровые образы отдельных приложений документа, которые не вшиты, а вложены в издание (альбомы схем, карты, графики и т. д.), лучше представить после основного массива цифровых образов оригинала.

При наличии в оригинале ярко выраженного просвечивания текста (иллюстраций) с оборотной стороны листа, приводящего к искажению или утрате информации, сканирование необходимо проводить с подкладыванием черного или иного однотонного листа бумаги, обеспечивающего нивелирование просвечивания при сохранении цветового фона страницы.

Если утрачена часть страницы, то она сканируется на фоне подложенного белого листа бумаги размером равного другим страницам документа.

Обработка электронных документов. Создание архивных и пользовательских копий

В результате сканирования создается два вида электронных копий документа: архивная (мастер-копия) и пользовательская (экранная) копия.

Архивная копия (мастер-копия) сохраняется в формате TIFF с максимальным качеством и с минимальной последующей обработкой⁵. Она предназначена для воссоздания максимально точной бумажной копии документа в случае его утраты, для полиграфических целей и как основа для изготовления других цифровых копий, в частности пользовательских. Данные копии необходимо размещать (сохранять) на носителях с длительным временем существования, например на оптических дисках (CD, DVD) или магнитных жестких дисках. Доступ к копиям данного типа ограничен, пользователям они не предоставляются или предоставляются только в крайнем случае.

Пользовательская (экранная) копия изготавливается из мастер-копии⁶. Она предназначена для просмотра на персональном компьютере локальными и удаленными пользователями. При создании пользовательской копии документа с помощью графического редактора сохраняется только часть страницы с текстом или изображением (без полей), устанавливается фиксированный размер страниц. Если информационная часть обложки, вклейки или иллюстрации занимает больше места, допускается изменение размера изображения, но в результате все страницы пользовательской копии должны быть одинакового размера. При обработке изображений допускается применение алгоритмов сжатия с потерями, но при этом должна сохраняться читабельность текста и возможность различения деталей графических элементов. Пользовательская копия сохраняется в формате PDF.

⁵ Порядок создания архивных копий на основе опыта работы АОНБ им. Н. А. Добролюбова приведен в Приложении № 4

⁶ Порядок создания пользовательских копий с применением программ Adobe Photoshop и Adobe Acrobat Pro приведен в Приложении № 5

Запись электронных копий на носители длительного хранения

Завершающим этапом в оцифровке документа является запись данных на оптические диски CD или DVD (в зависимости от объема информации). На диск пишется папка с архивными копиями (мастер-копиями) и папка с пользовательскими копиями (или пользовательский файл в формате PDF). В итоге в корне диска будет находиться папка с номером документа, например папка 17кр001, внутри нее — папка arhiv с архивными файлами и папка user с пользовательскими копиями (или файл 07кр001.pdf). Название диска присваивается в соответствии с идентификатором документа, например 07кр001. Желательно, чтобы по окончании работ на жестком диске компьютера или на съемном жестком диске сохранялся тот же набор папок и файлов.

Правила формирования идентификационного номера электронного документа

Идентификационный номер электронного документа состоит из 3 блоков:

- первый блок обозначает год оцифровки документа и состоит из 2 цифр;
- второй блок обозначает вид оцифрованного документа и состоит из 2 латинских букв;
- третий блок обозначает порядковый номер оцифрованного документа текущего года и состоит из 3-х цифр.

Например, 17kr001.

Обозначения видов документов:

Книги:

- kr — книги печатные;
- kg — книги рукописные;
- kd — диссертации, авторефераты;
- ka — альбомы (изоиздания);
- kn — ноты;
- kk — карты (атласы);
- ks — статьи из сборников.

Периодические издания:

- pg — газеты;
- pz — журналы;
- ps — статьи из периодических изданий.

Листовые материалы:

- li — изобразительные материалы;
- lp — листки (печатные);
- lr — рукописи;
- ln — ноты;
- lk — карты.

Другое (документы, созданные работниками библиотеки):

- du — библиографические указатели, списки;
- dk — каталоги выставок;
- dp — презентации.

Бланк сопровождения документа

(лицевая сторона)

Идентификатор							
Автор							
Заглавие							
Место хранения						инв. №	
Кол-во стр.							
Режим сканир-я						dpi:	
Цель оцифровки документа:							
сохранность	докомплектование			другое			
Коллекция:							
Электронное издание:							
Диски:	CD			DVD			
Особенности сканирования документа:							
Переплет	да			нет			
Форзац	да			нет			
Шмуцтитул	да			нет			
Оборот шмуцт.	да			нет			
Титул. лист.	да			нет			
Оборот тит. л.	да			нет			
Иллюстрации							
Карты							
Штампы, автогр.							
Другое							

(оборотная сторона)

Ф.И.О. сканировщика:							
Сканер							
Цвет скан.	серый	ч/бел	цветн.	смеш.			
Кач. копии	архив.			архив. выс. разр.			
	низкого разр.			копия другого типа			
	низкого разр. (pdf)						
Кол-во файлов							
Объем TIFF Мб:							
Объем PDF Мб:							
Формат сжатия							
Вид эл. ресурса:	изображ.		текст		другое		
Кол-во файловых форм:	один формат файлов			несколько форматов файлов			
Гарантии качества копии:	присутствует			неизвестно (чужая копия)			
Предшественник копии:	ори-гинал	м/ф	комп. файл	сме-шан	неиз-вестно		
Цена эл. копии							
<i>Имя варианта настройки</i>							
<i>Нумерация с</i>							
<i>Размер кадра</i>							
Примечания							

Порядок создания архивных копий документов

1. Создать на жестком диске компьютера папку с номером идентификатора, например 17кр001.
2. Расположить документ на рабочей плоскости сканера.
3. Осуществить подбор режима сканирования в окне настройки сканера — разрешение, формат расположения документа, режим серого и т. д. После предварительного (пробного) сканирования подобранные параметры сохраняются в настройках сканера.
4. Подшивка газет расширяется, если в собранном варианте невозможно отсканировать страницу полностью.
5. Итогом сканирования должно стать изображение полной страницы с небольшим (1 см) полем и областью переплета.
6. Создать в папке с номером идентификатора документа вложенную папку arhiv, например 17кр001/arhiv.
7. Открыть файл, подлежащий обработке. Если требуется, повернуть изображение по часовой стрелке или против неё, используя в пункте меню *Изображение* опцию *Поворот холста*.
8. При необходимости использовать режим обрезки (меню *Изображение* — *Обрезать*): в случае, если поле и область переплета превышает 1 см.
9. Коррекция изображения для создания архивной копии не производится.
10. Сохранить каждое изображение отдельным файлом в папке 17кр001\arhiv с именем, сформированным по присвоенному документу идентификатору с добавлением в конце номера страницы через символ подчеркивания, например 17кр001_01.tif. Изображение сохраняется без компрессии.
11. Внести информацию о количестве созданных файлов в бланк сопровождения документа.

Порядок создания пользовательских копий документов

1. Создать в папке с номером идентификатора документа вложенную папку user — 17кр001/user.
2. В программе **Adobe Photoshop** открыть файл, подлежащий обработке, из папки arhiv — 17кр001/arhiv; при необходимости выровнять изображение. Выравнивание производится по первой строке или по левому краю текста.
3. При обработке книг по самому широкому изображению печатного поля обычного листа выставить рамку фиксированного размера. Для этого использовать инструмент *Прямоугольное выделение* в *Панели инструментов программ*. Стил: *Фиксированный размер*. Размер рамки выставляется в пикселях и подгоняется так, чтобы в выделенной области был только печатный текст или изображение, допускается поле в несколько десятков пикселей. Рамка будет использоваться для обработки всего документа. Если документ будет обрабатываться в течение нескольких рабочих дней, параметры рамки следует зафиксировать.
4. Осуществить обрезку изображения.
5. Обработать изображения, используя настройки яркости, контрастности, выравнивание уровня серого и т. д.
6. Сохранить файл изображения в папке 17кр001/user с именем, присвоенным архивной копии данной страницы, например 17кр001_01.jpg.

Сохранение файлов с использованием программы Adobe Acrobat Pro

1. Открыть Adobe Acrobat Pro и выполнить команду *Файл — Создать — Объединить файлы в один документ PDF* или *Начало работы — Объединить файлы в портфолио PDF*. Далее выполнить команду *Добавить файлы*. Для загрузки в программу выделить все изображения из папки 17кр001/user. Проверить, чтобы файлы были сортированы по имени, а порядок страниц был верным, и нажать кнопку *Объединить файлы*.
2. Сохранить пользовательскую копию документа (*Файл — Сохранить как — PDF*) в папке с именем идентификатора, указанным на сопроводительном бланке, например 17кр001.pdf.
3. Для распознавания PDF-файла выбрать в правой стороне pdf-документа панель инструментов, далее *Инструменты — Распознавание текста — В этом файле*. Убедиться, что для распознавания выбраны все страницы, язык распознавания — русский.
4. Для создания закладок в PDF-файле открыть слева кнопку *Закладки*. Используя инструмент *Выделение*, выделить область страницы, которую необходимо отметить закладкой. Выделить текст мышью, чтобы отметить его закладкой. Выделенный текст становится меткой новой закладки. Эту метку можно редактировать. Щелкнуть закладку, после которой нужно поместить новую закладку. Если закладка не выделена, то новая закладка автоматически добавится в конец списка. На панели «Закладки» ввести или изменить имя новой закладки.
5. Иерархия созданных в pdf-документе закладок должна полностью соответствовать оглавлению (содержанию) оригинала.
6. Перед сохранением и закрытием файла проверить корректность созданных закладок (орфографию, пунктуацию) и правильность навигации по ним.