**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО СКАНИРОВАНИЮ**

**И ОБРАБОТКЕ ЭЛЕКТРОННЫХ КОПИЙ ДОКУМЕНТОВ**

1. СОЗДАНИЕ ЦИФРОВЫХ КОПИЙ
   1. **Получение документа для оцифровки (сканирования)**

Документ, предназначенный для оцифровки, оператор по сканированию получает в отделах-фондодержателях вместе с дефектными ведомостями на печатный оригинал.

* 1. **Подготовка к сканированию**
     1. В начале каждой смены и по мере загрязнения, оператор должен протереть от пыли и отпечатков рук (пальцев) салфеткой рабочую поверхность сканера.
     2. Исключить расположение светильников общего освещения непосредственно над сканером, для того, чтобы исключить отражение света от листа, которое не должно попадать в камеру. Добиться, чтобы освещение всей области сканирования было равномерным, исключающим тени, блики, области повышенной засветки от ламп
     3. Перед началом сканирования оператору необходимо проверить настройки ПК, сканера и произвести пробное сканирование страницы документа, чтобы исключить возможные дефекты сканирования
     4. Создать в папке «Для оцифровки» папку с именем документа. В нее будут помещены файлы с изображениями страниц объекта, полученные в процессе сканирования.
     5. Отсканированные документы сохраняются в формате TIFF.
  2. Присвоение имен файлам с изображениями страниц
     1. Каждый файл электронной копии должен иметь уникальное имя. В основу имени ложится регистрационный номер, присвоенный электронной копии документа (ЭКД). (См. «Инструкцию по присвоению имен файлам»)
     2. К каждой полученной копии оператором добавляются характерные для неё идентификационные данные. А именно: блок сведений о порядковом номере страницы

2. ФОРМИРОВАНИЕ И ОБРАБОТКА МАССИВОВ ЦИФРОВЫХ КОПИЙ

2.1. Структура электронного информационного ресурса

В результате проведения работ по сканированию и обработке должен быть создан электронный информационный ресурс (ЭИР) представляющий из себя три массива цифровых копий документа:

* массив архивных копий (в формате TIFF)
* массив пользовательских копий (в формате JPEG)
* массив просмотровых копий (в формате PDF)

2.2. Обработка и корректировка массивов цифровых копий

Требования к сканированию и обработке полученных массивов цифровых копий разработаны на основе «Критериев качества цифровых копий, обработанных графических образов и многостраничных файлов в формате PDF».

В таблице № 1 приведены форматы файлов, рекомендуемое программное обеспечение и методы сжатия файлов.

Таблица № 1

|  |  |  |
| --- | --- | --- |
| Название формата и цель использования | Программы, которые могут открывать файлы | Рекомендуемый метод сжатия |
| TIFF для мастер-копий | Большинство растровых редакторов и настольных издательских систем; векторные редакторы, поддерживающие растровые объекты | Сжатие LZW (без потери качества) |
| JPEG для  пользовательских и просмотровых копий | Большинство графических программ, интернет-браузеры | Степень сжатия с сохранением 80% качества исходного файла |
| PDF для  пользовательских и просмотровых копий | Межплатформенный формат файла, переносимый на компьютеры любых систем. Для просмотра файла нужна программа Adobe Reader или аналог, в т.ч. плагины браузера | Оптимизация для WEB, сжатие JPEG до 50%, понижение разрешения до 150 DPI |

2.2.1. Создание архивных копий документов (мастер-копий)

Для создания архивной копии используются файлы, полученные при сканировании документа. Обработка проводится в программе FineReader или любой другой графической программе, установленной на компьютере. Создание и сохранение файлов архивных копий производится в формате TIFF.

Требования к полученным архивным копиям документа

* 1. Результатом сканирования издания должны быть графические образы в формате TIFF в постраничном представлении со сквозной нумерацией файлов-имиджей страниц (согласно «Инструкции по присвоению имен файлам»).
  2. При сканировании должны захватываться границы документа с запасом не менее 3-5 мм, но не более 1 см.
  3. Графические образы должны полностью соответствовать страницам оригинала издания. Никакой коррекции изображения не производится.
  4. Сканирование газет должно производиться в сером режиме Grayscale 8 Bit (256 градаций серого цвета), с настройками яркости и контрастности, выставленными для оптимального зрительного восприятия и эффективного распознавания типографского текста.

Страницы, содержащие цветные, черно-белые или серые иллюстрации, должны сканироваться в цветном режиме (RGB) - 24-bit color с разрешением 300 DPI.

* 1. Все элементы одного издания должны быть отсканированы с одинаковым разрешением. Минимальные требования к оптическому разрешению: не ниже 300 DPI
  2. Издания сканируются полностью, включая титульный лист и оборот титульного листа, концевой титульный лист (колофон), 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.
  3. При наличии в оригинале ярко выраженного просвечивания текста (иллюстраций) с оборотной стороны листа, приводящих к искажению или утрате информации, сканирование должно проводиться с подкладыванием черного или иного однотонного листа бумаги, обеспечивающего нивелирование просвечивания.
  4. Если утрачена площадь страницы (отсутствие части листа по краю либо в любой другой части страницы), то страница сканируется на фоне подложенного белого листа бумаги размером равного другим страницам документа для закрытия последующей либо предыдущей страницы издания.
  5. Сканирование должно производиться при расправленных страницах (замятия, загибы страниц) для максимально полного раскрытия страницы и текста документа.
  6. Ориентация графических образов должна соответствовать ориентации оригинала.
  7. Не допускается наличие волнообразных страниц.
     1. Графические образы должны быть четко сфокусированы (резкость текстовых графических образов должна фокусироваться по тексту). Оттенки, глубина и насыщенность цвета образов должны быть максимально приближены к оригиналу и максимально единообразны в пределах одной книги.
     2. Все образы страниц одного формата должны быть одинакового размера по высоте и ширине. Образы большеформатных или мелкоформатных страниц, отличных по размеру от основной части страниц источника сканирования, должны сохранять оригинальный размер.
     3. Если разворот или вкладка имеет размер, превышающий поле сканирования используемого оборудования, то допускается оцифровка отдельными частями. Полученные изображения частей должны быть соединены в одно изображение. В ряде случаев допускается сохранение изображений частями, если эти части не являются логически - цельной информацией (например, несколько различных таблиц на одном большом листе - вложении могут быть сохранены раздельно).
     4. Вклейка, отличающаяся по одному из параметров от размеров основного массива страниц документа, сканируется на фоне предшествующей или следующей за вклейкой страницы без подкладывания посторонних листов.
     5. Цифровые образы отдельных приложений, которые не вшиты, а вложены в издание (альбомы схем, карты, графики и т.д.) должны быть представлены после основного массива цифровых образов оригинала.
     6. Не допускается сохранение нескольких изображений в одном графическом файле, съемка с разным разрешением в пределах одного объекта.
     7. Изображение сохраняется без компрессии.

2.2.2. Создание пользовательских копий

Пользовательская копия изготавливается из архивной копии документа. Сохранение файлов пользовательских копий производится в формате JPEG.

Создается в программе FineReader или любой другой графической программе, установленной на компьютере. При создании нужно учитывать качество изображений и стремиться к наименьшему возможному размеру файла.

Требования к обработанным копиям изданий в формат JPEG

* + - 1. Графические образы из формата TIFF в формат JPEG должны быть переконвертированы.
      2. Обработка графических образов не должна приводить к существенным изменениям получаемых образов по сравнению с оригиналом.
      3. Графические образы должны полностью соответствовать страницам оригинала, потеря информации или включение отсутствующей в оригинале информации не допускаются.
      4. При обработке не допускается обрезка полей страниц издания, а также удаление фрагментов (текста, изображений, пометок, печатей и т.п.). При обработке не допускается зарезка текста. Размер полей должен максимально соответствовать размеру полей печатного оригинала.
      5. Графический образ страницы должен быть выровнен по горизонтали по верхнему краю страницы печатного оригинала.
      6. Для страниц одного формата графические образы должны иметь одинаковый геометрический размер в пикселях. Цифровые образы большеформатных или мелкоформатных страниц, отличных по размеру от основной части страниц оригинала, должны сохранять оригинальный размер и соотношение размеров с основным массивом страниц.
      7. Не допускается наличие по краям образов черных (белых) полей, фрагментов других страниц или составных частей оригинала.
      8. Развороты вшитых\вклеенных в оригинал листов, на которых расположена логически цельная информация (распашная карта/ изображение, несколько изображений под одним названием, единые таблицы и т.д.), должны быть представлены как один цифровой образ с размером, отличающимся от размера основных страниц Издания.
      9. Развороты вшитых\ вклеенных в оригинал листов, информация на которых логически не связана, должны быть представлены как два цифровых образа.
      10. Недопустимы заломы строк на образе, характерные для «тугого» переплета.

11.Имена файлов-имиджей страниц должны соответствовать «Инструкции по присвоению имен файлам»

**2.2.3.** Создание многостраничных файлов формате PDF

Результатом заключительного этапа оцифровки является создание многостраничного файла в формате PDF/А. Изготавливается из пользовательской копии. Допускается применение алгоритмов сжатия с потерями, уменьшение разрешения графического образа, внедрение в документ дополнительных данных (например, водяных знаков).

Требования к созданию многостраничных PDF-файлов

* + - * 1. Электронные копии документов представляются в формате *PDF/А* в соответствии со спецификацией *PDF/A-1b* в виде многостраничного электронного документа с разрешением не менее 200 dpi.
        2. При формировании PDF-файлов необходимо использовать способ сжатия графических образов, позволяющий оптимизировать файлы по размеру и качеству.
        3. Для возможностей контекстного поиска в каждом PDF-файле должен быть сформирован так называемый подслой автоматически распознанного текста с привязкой по страницам.
        4. Должен соблюдаться единый размер графических образов в рамках одного документа: все графические образы страниц одного формата должны быть одинакового размера по высоте и одного размера по ширине.
        5. Графические образы большеформатных или мелкоформатных страниц, отличных по размеру от основной части страниц документа, должны сохранять оригинальный размер.

3. КОНТРОЛЬ КАЧЕСТВА ЭЛЕКТРОННЫХ КОПИЙ

Контроль качества электронных копий производится систематически, на разных этапах создания ЭИР.

1. На этапе сканирования и графической обработки документа: самопроверка и самоконтроль осуществляется оператором сканирования.
2. По окончании формирования массивов электронных копий: ответственным сотрудником проводится контрольная проверка на соответствие «Критериям качества цифровых копий, обработанных графических образов и многостраничных файлов в формате PDF». Наличие искажений, потерь информации и других особенностей оригинала, повлиявших на качество электронной копии, а так же дефекты сканирования, выявленные в ходе этой проверки должны быть зафиксированы в дефектной ведомости (см. «Инструкцию по заполнению дефектной ведомости на ЭКД»).