

«Единые технические требования к оцифровке
изданий, включаемых в НЭБ»

Классификация документов библиотечных фондов

Библиотечные фонды по отношению к подходам по их оцифровке условно разбиваются на 3 категории:

- I категория "Раритеты": книжные памятники (до 1830 г.), особо ценные и редкие книги, рукописи.
- II категория "Художественно-культурные ценности и социально значимые издания": издания, представляющие особую художественную и культурную ценность, содержащие авторские иллюстрации и рисунки, а также издания, имеющие большую социальную значимость.
- III категория "Информационно значимые издания": тиражная литература, которая представляет ценность только как информация, а не как оригинальное издание.

Указанные ниже рекомендации и технические требования являются минимальным уровнем для оцифровки изданий, включаемых в НЭБ. Фондодержатель может самостоятельно повысить уровень требований к оцифровке.

Рекомендации к составу оборудования

При выполнении работ по сканированию могут быть использованы следующие типы специализированного оборудования:

- комплексы высококачественного сканирования – применяются для I и II категорий фондов. Использование прижимного стекла не допускается.
- специализированные книжные сканеры с использованием специальных колыбелей (ложементов) для документов во избежание изгиба переплета – применяются для II и III категорий фондов. Допускается использование прижимного стекла в случае удовлетворительного состояния издания. Если угол раскрытия книги ограничен, ее сканирование производится только с применением специальной колыбели (ложемента), обеспечивающей малый угол раскрытия изданий (до 120°)
- специализированные поточные высокоскоростные сканеры – в особых случаях по решению владельца фонда допускается только для III категории фондов.
- планшетные документные сканеры - в особых случаях по решению владельца фонда допускается только для III категории фондов.
- специализированные сканеры микроформ, сканеры слайдов с высоким разрешением – применяются для материалов на пленках.

Рекомендуется использовать сканирующее оборудование обеспечивающее читаемость предельной группы с пространственной частотой линий 5,6 тест-объекта разрешения, выполненного по ГОСТ 13.1.701-95 (выдержка из ГОСТ – группу элементов мира линий считают читаемой, если оба составляющих ее элемента не сливаются с фоном по всей длине). Для измерения читаемости предельной группы, перед началом работ сканируется специальная мишень, определённая ГОСТ 13.1.701-95 и та группа линий, части которой (белые и черные линии) не сливаются, считается читаемой.

Требования к сохранности документного фонда при оцифровке

Главное требование к процессу оцифровки заключается в обеспечении максимальной сохранности первоисточников (оригиналов) и недопущении воздействий, ведущих к ухудшению их физического состояния в процессе выполнения работ.

С целью обеспечения сохранности изданий необходимо на этапе подготовки списков в фондах выявлять издания, которые по своему физическому состоянию не могут пройти сканирование без повреждений. Такие издания должны быть исключены из планов оцифровки до момента приведения их в состояние, пригодное для сканирования.

Требования к оригиналам для прохождения сканирования бесконтактным способом на специализированных книжных сканерах и комплексах высококачественного сканирования для всех категорий фондов:

а) Состояние бумаги допускает перелистывание всех страниц. Если издание подверглось негативному воздействию (окислению, пересушиванию и т.п.) и бумага стала хрупкой (отламывается при сгибании), то рекомендуется предварительно провести нейтрализацию.

б) Все страницы издания должны быть отделены друг от друга и позволять постранично перелистывать издание. Если отдельные страницы не были разрезаны в типографии или были некорректно вшиты, то их необходимо разделить перед сканированием.

в) Издание можно раскрыть на угол более 120 градусов. Если при раскрытии издания возникает угроза повреждения корешка, то необходимо предварительно его отреставрировать.

г) Переплет позволяет при открытии издания считать всю информацию. Если часть информации «уходит» в шивку переплета и возникает значительный изгиб листов, то при оцифровке потери информации усилятся из-за теней и с изображения можно будет считать только часть информации. Для выполнения сканирования желательно заменить такое

издание на другой экземпляр или изменить его переплет для считывания полного информационного содержания со всех листов.

При соответствии издания указанным критериям в результате проведения работ по оцифровке процент повреждения страниц документов-первоисточников не должен превышать 0,01% (1 поврежденная страница на 10 000 страниц) за исключением рукописей. Повреждение страниц документов-первоисточников рукописей не допускается.

Все операции с изданиями должны производиться очень аккуратно, при соблюдении следующих правил:

- хранить издания необходимо только в горизонтальном положении;
- запрещается держать издания на весу;
- запрещается делать пометки на страницах изданий и использовать другие способы маркировки документов, приводящие к изменению изначального состояния издания.

Перемещение книжных материалов должно производиться, не допуская разрушение и повреждение документов (отсутствие агрессивных компонентов (лигнин и др.) для бумаги документов в оборудовании, применяемом для перемещений).

При перемещении вне помещения, в холодное время, обязательным требованием является проведение акклиматизации книг для исключения образования конденсата.

Температура, относительная влажность воздуха, интенсивность освещения и состав света, пыль и загрязнение воздуха в помещениях должны соответствовать требованиям ГОСТ 5.50-2002.

Не допускается механическое воздействие на красочный слой страниц первоисточника при сканировании.

Для III категории фондов в особых случаях допускается сканирование контактным способом на планшетных или поточных протяжных сканерах.

Применение планшетных сканеров возможно для оригиналов, соответствующих следующим критериям:

- а) поверхность оцифровки плоская или может быть распрямлена без ущерба для объекта,
- б) поверхность оцифровки не может быть повреждена при контакте со стеклом,
- в) для специальных видов объектов, требующих максимального разрешения (более 1000 DPI, например, слайдов),

г) свет устройства не наносит вреда объекту сканирования.

Применение специализированных поточных протяжных сканеров допустимо по решению владельца фонда. Это возможно в случае сканирования большого количества документов, представленных отдельными листами (например, каталоги). В случаях, когда позволяет состояние материала (новые издания) и по решению владельца фонда, может быть выполнена расшивка переплетенных изданий для сканирования на поточных сканерах. При этом порядок расшивки, а также последующего возврата издания владельцу должен быть предварительно согласован исполнителем работ по оцифровке.

Требования к источникам света, применяемым для проведения сканирования

В качестве источника освещения должна использоваться «холодная» электролюминесцентная подсветка. Основным требованием к освещению является его безопасность для оригинала, обеспечение равномерности освещения по формату оригинала для получения качественных результатов и сохранения естественного свето-теневого баланса.

Для освещения следует использовать флуоресцентные лампы с фильтром, защищающим от ультрафиолетового излучения и поглощающим тепло, или волоконно-оптические системы освещения.

Осветители, которые не наносят вред объекту сканирования потоком ИК и УФ излучения, должны соответствовать следующим требованиям:

- Норма освещенности на поверхности документов при экспонировании и в момент осмотра – не более 150 лк;
- Источники света должны обеспечивать оптическое излучение, длина волны которого не менее 400 и не более 760 нм;

Выбор между типами осветителей осуществляется совместно с хранителем объекта и специалистом по консервации.

Требования к сканированию при создании мастер-копий

Для всех видов изданий, относящихся к I и II категориям библиотечных фондов сканирование всех элементов производится в соответствии с требованиями к созданию мастер-копий. Для фондов III категории сканирование с параметрами мастер-копии не предусмотрено.

В результате сканирования изданий-первоисточников должны быть получены графические образы в формате TIFF со сквозной нумерацией файлов-имиджей страниц. С целью недопущения потери качества изображений примеряется алгоритм сжатия только LZW-compressed.

Цветовой режим RGB Color 24 Bit (воспроизведение до 16,7 миллиона цветов на пиксел).

В процессе сканирования каждого издания необходимо включение в изображение сертифицированной цветовой мишени. Правила работы с цветовыми мишенями должны быть разработаны дополнительно.

При сканировании захватываются границы издания и цветовой мишени с запасом не менее 3-10 мм.

Все элементы одного издания должны быть отсканированы с одинаковым разрешением. Минимальные требования к оптическому разрешению:

- 1) Для фондов I категории: не ниже 600 DPI
- 2) Для фондов II категории не ниже 300 DPI

Требования к сканированию при создании пользовательской копии

Для всех видов изданий, относящихся к III категории библиотечных фондов сканирование всех элементов производится в соответствии с требованиями к созданию пользовательских копий.

В результате сканирования изданий-первоисточников должны быть получены графические образы в формате JPEG с сохранением 80% качества при сжатии со сквозной нумерацией файлов-имиджей страниц.

Возможны 2 подхода к выбору цветового режима:

1. Использование RGB Color 24 Bit (цветного) для всех страниц издания. Рекомендуется для полноцветных иллюстрированных изданий.

2. Использование Grayscale 8 Bit (256 градаций серого цвета) для основной части издания и режима RGB Color 24 Bit (цветного) для отдельных элементов, Рекомендуется в основном для изданий, отпечатанных с использованием одноцветной печати:

- А) Режим Grayscale 8 Bit (256 градаций серого цвета) применяется для:
- страниц, содержащих текст,

- страниц со штриховыми монохромными иллюстрациями,
- вакатов, включая оборотные стороны любых иллюстраций, в том числе с печатями, штампами, карандашными и т.д. пометами, привнесенными после выпуска издания (не принадлежащих изначально оригиналу).

Б) Режим 24-bit RGB (цветной режим) применяется для:

- обложки / суперобложки (лицевой и оборотной стороны),
- части форзацев (переднего и заднего), наклеенные на оборотную сторону обложки,
- страницы и приложения, содержащие цветные иллюстрации или иллюстрации в градациях серого цвета,
- страницы, имеющие авторские или владельческие пометы (посвящения, автографы).

При сканировании захватываются границы листа с запасом 1-5 мм.

Все элементы одного издания должны быть отсканированы с одинаковым разрешением. Минимальные требования к оптическому разрешению: не ниже 300 DPI

Требования к обработке мастер-копии

Для мастер-копии проводятся следующие виды обработки изображений:

- 1) Сшивка частей изображения для широкоформатных оригиналов, которые не было возможности отсканировать целиком.
- 2) Обрезка избыточного фона сканирования. Допустимый размер полей 3-10 мм от края издания и цветовой мишени.

Требования к обработке пользовательской копии для фондов I категории

Для фондов I категории пользовательская копия создается из файлов мастер-копии путем конвертации файлов TIFF в формат JPEG с сохранением 80% качества при сжатии и выполнении обработки полученных файлов.

- 1) Значение разрешения не менее 300 DPI.
- 2) Обрезка проводится с полями 3-10 мм от края издания включая обложку и каскад страниц. Изображение цветовой мишени отсутствует.

3) Для страниц одного формата графические образы должны иметь одинаковый размер в пикселях по высоте и ширине. Единый размер должен быть выбран по наибольшему изображению страницы с запасом не более 1-3 мм, при этом для некоторых страниц поля с фоном могут добавляться в необходимом объеме до размеров наибольшего изображения страницы документа с целью сохранения пропорций.

Требования к обработке пользовательской копии для II категории фондов

Для фондов II категории пользовательская копия создается из файлов мастер-копии путем конвертации файлов TIFF в формат JPEG с сохранением 80% качества при сжатии и выполнении обработки полученных файлов:

- 1) Значение разрешения не менее 300 DPI.
- 2) Обрезка проводится с полями 3-10 мм от края издания включая обложку и каскад страниц. Изображение цветовой мишени отсутствует.
- 3) Приведение к единому размеру графических образов в пикселях по высоте и ширине в рамках одного издания для страниц одного формата.

Требования к обработке пользовательской копии для III категории фондов

Для фондов III категории пользовательская копия создается на этапе сканирования и сразу представлена файлами формата JPEG с сохранением 80% качества при сжатии. Последующая обработка полученных файлов включает в себя следующие операции:

- 1) Обрезка по краю листа. Изображение краев обложки и каскада страниц отсутствует.
- 2) Приведение к единому размеру графических образов в пикселях по высоте и ширине в рамках одного издания для страниц одного формата.

Требования к созданию просмотровой копии

В соответствии с целями создания просмотровой копии необходимо обеспечить удобство просмотра в Интернет, то есть собрать отдельные изображения в многостраничный файл, уменьшить объем этого файла, обеспечить возможность полнотекстового поиска.

Для всех категорий фондов требования к созданию просмотровой копии одинаковые:

- 1) Просмотровая копия создается из файлов JPEG, прошедших обработку при создании пользовательской копии.
- 2) Просмотровая копия должна быть представлена в виде файла формата PDF.
Требования к PDF:

- должен быть оптимизирован для WEB;
- должен содержать подложку из результатов автоматического распознавания текста без верификации;
- цветовой режим изображений не изменяется;
- возможно понижение разрешения до 150 DPI;
- возможно дополнительное сжатие изображений, но не более чем 50%;

Требования к оцифровке различных видов копий собраны в таблице 1.

Таблица 1 Требования к оцифровке различных видов копий.

Фонды и копии	Мастер-копия	Пользовательская копия	Просмотровая копия
I категория "Раритеты" :	<ul style="list-style-type: none"> – Сканирование обложки и всех страниц без пропусков, – Не подлежит обработке и выдаче пользователям, – TIFF без потери качества (LZW), – RGB Color 24 Bit, – Сертифицированная цветовая мишень, – Разрешение: от 600 DPI, – Поля 3-10 мм от образов края издания и цветовой мишени. 	<ul style="list-style-type: none"> – Создание на основе мастер-копии для I категории, – JPEG, сохранение 80% качества, – RGB Color 24 Bit, – Разрешение: от 300 DPI, – Поля 3-10 мм от образа края издания (без образа цветовой мишени), – Для страниц одного формата графические образы должны иметь одинаковый размер в пикселях по высоте и ширине в рамках одного издания 	<ul style="list-style-type: none"> – Создание на основе пользовательской копии для I категории, – PDF, максимально допустимое сжатие 50%, – Автоматическое распознавание без верификации, – RGB Color 24 Bit, – Разрешение: от 150 DPI.
II категория "Художественно-культурные ценности" :	<ul style="list-style-type: none"> – Сканирование обложки и всех страниц без пропусков, – Не подлежит обработке и выдаче пользователям, – TIFF без потери качества (LZW) – RGB Color 24 Bit, 	<ul style="list-style-type: none"> – Создание на основе мастер-копии для II категории, – JPEG, сохранение 80% качества, – RGB Color 24 Bit, – Разрешение: от 300 DPI, – Поля 3-10 мм от образа края издания (без образа цветовой мишени), 	<ul style="list-style-type: none"> – Создание на основе пользовательской копии для II категории фондов, – PDF, максимально допустимое сжатие 50%, – Автоматическое распознавание без

	<ul style="list-style-type: none"> – Сертифицированная цветовая мишень, – Разрешение: от 300 DPI – Поля 3-10 мм от образов края издания и цветовой мишени. 	<ul style="list-style-type: none"> – Для страниц одного формата графические образы должны иметь одинаковый размер в пикселях по высоте и ширине в рамках одного издания, – Выравнивание по верхнему краю страницы, – Выравнивание строк проводится исходя из требования соответствия изображения оригиналу и недопустимости потерь информации. 	<ul style="list-style-type: none"> верификации, – RGB Color 24 Bit, – Разрешение: от 150 DPI.
<p style="text-align: center;">III категория "Информационно значимые издания":</p>	<ul style="list-style-type: none"> – Не создается. 	<ul style="list-style-type: none"> – Сканирование обложки и всех страниц без пропусков, – JPEG, сохранение 80% качества, – Greyscale 8 Bit (текст) и RGB Color 24 Bit (цветные иллюстрации), – Разрешение: от 300 DPI, – Обрезка по краю листа, – Для страниц одного формата графические образы должны иметь одинаковый размер в пикселях по высоте и ширине в рамках одного издания, – Выравнивание по верхнему краю страницы, – Выравнивание строк проводится исходя из требования соответствия изображения оригиналу и недопустимости потерь информации. 	<ul style="list-style-type: none"> – Создание на основе пользовательской копии для III категории фондов, – PDF, максимально допустимое сжатие 50%, – Автоматическое распознавание без верификации, – Greyscale 8 Bit (текст) и RGB Color 24 Bit (цветные иллюстрации), – Разрешение: от 150 DPI.