Файлы в библиографических записях

Обложки и страницы изданий в записях каталога — публикация и хранение

Основные понятия
Недостатки существующих систем хранения
Как это устроено
Как файлы попадают в запись
- Какие файлы идут в запись
- Обложка или прочий файл
- Строка ссылки в каталоге
- Повторная передача

Публикация файлов
- Когда файл уходит в архив
- Состояние файла
- Обложки и страницы — разного размера
- Замена файла и смена типа

Как размещается файловый архив
Возможности файлового архива
- Основные функции
- Расширяя по запросу от библиотек

Частые вопросы

Каталогизатор не только заполняет поля записи. К записи можно приложить файлы: обложку, титульный лист, оглавление, сканы страниц, аудиоматериалы. В запись уходит не сам файл, а ссылка на него. Файл лежит в файловом архиве библиотеки и открывается по этой ссылке из каталога библиотеки.

Страница для каталогизаторов и администраторов АБИС. Здесь описано, как ИИ-каталогизация публикует файлы. Какие файлы попадают в запись и в каком виде, когда файл публикуется и как устроено хранение.

Основные понятия

- **Файл записи** — снимок, который сотрудник отметил галочкой «В запись».
- **Обложка** и **прочий файл** — два вида файлов записи. Вид зависит от типа файла. Обложка открывается небольшой картинкой в каталоге, прочие файлы — крупно.
- **Строка ссылки** — поле записи с адресом файла. Адрес стоит в поле 856 в RUSMARC и MARC 21, в поле 951 в ИРБИС.

- **Файловый архив библиотеки** — сервис, разработанный нашей организацией, который хранит файлы записей и отдаёт их по ссылке. Адрес — files.bibtech.ru. Он работает отдельно от сервиса обработки.
- **Публикация** — отправка файла в файловый архив. После неё ссылка в записи открывается.
- **Сеть библиотек** — библиотеки с общей настройкой сервиса, например библиотеки одной области или одной ЦБС.

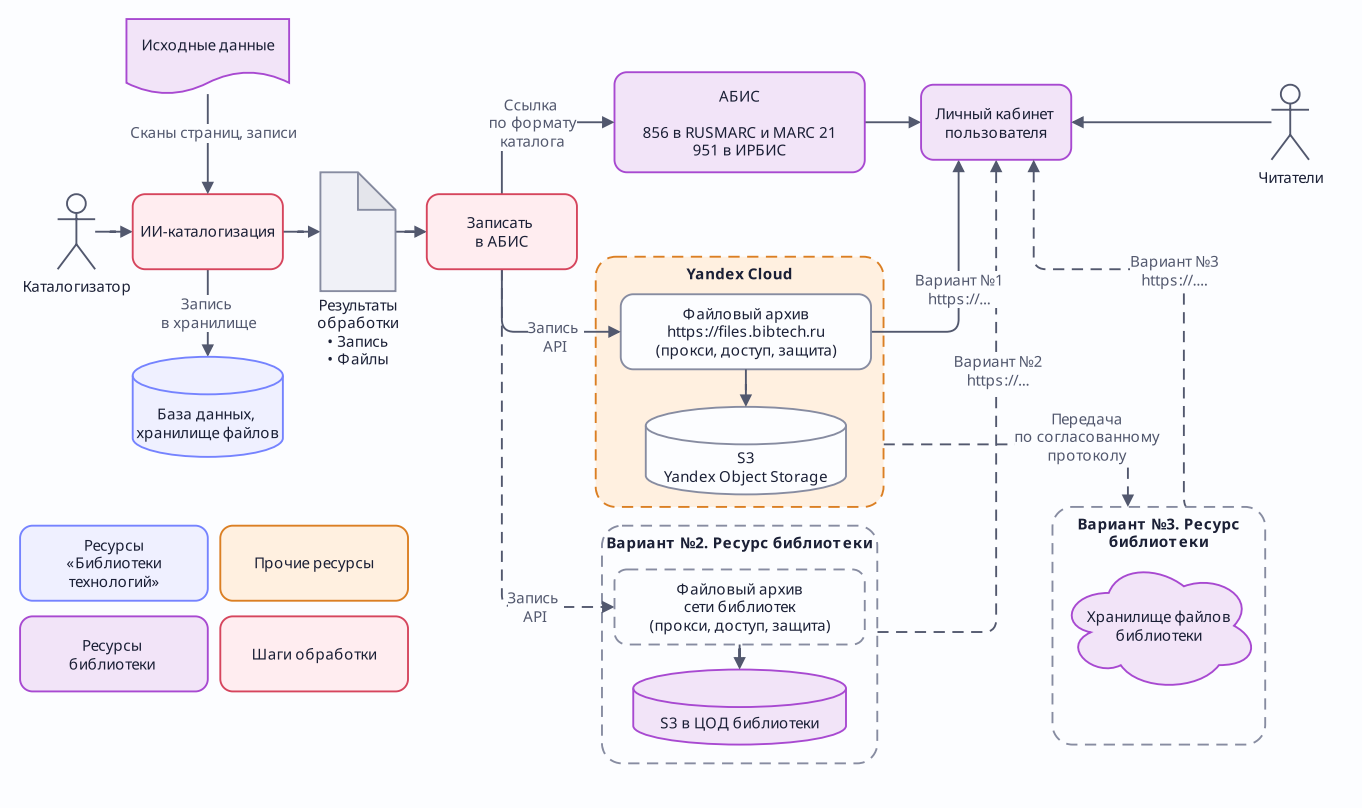
Недостатки существующих систем хранения

Есть два основных способа хранения файлов в записи: - в АБИС, в папке на сервере библиотеки; - показывают по ссылке.

Как правило в хранении и доступе есть следующие недостатки:

- **Файл отдаётся как есть.** Один и тот же файл показывается и в выдаче, и в карточке записи, и там, где нужно прочитать оглавление. Маленький файл плохо виден, когда нужно разобрать текст. Большой браузер уменьшает сам: страница грузится долго, интерфейс разъезжается, канал библиотеки и читателя нагружен. Большой файл в запись и не положить. Разработчики ИРБИС64 советуют держать изображения внутри записи «ориентировочно не более 50 кб». В Koha обложки лежат в базе данных. Участники сообщества просят хранить их вне базы и отмечают, что обработка обложек без нужды увеличивает размер базы.
- **Обработка и размещение — вручную.** Снимок нужно обрезать, поставить водяной знак, сделать несколько размеров: миниатюру для выдачи и крупный для чтения. Это работа в графическом редакторе для каждого файла. Каждый размер — отдельный файл, его тоже надо разместить, хранить и не потерять.
- **Чужой файл могут закрыть или подменить.** Если показывать файл с чужого сайта, нет гарантии, что владелец не закроет доступ и не заменит содержимое. При этом файл показывается на сайте библиотеки, от её имени. По данным Pew Research Center (https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears/) (2024), 38 % веб-страниц, которые были в сети в 2013 году, к 2023 году недоступны. Внешние сервисы обложек ещё и ограничивают показ: Open Library (https://openlibrary.org/dev/docs/api/covers) по ISBN отдаёт не больше 100 обложек с одного адреса за 5 минут, дальше отвечает ошибкой 403.
- **Файл привязан к серверу и АБИС.** В ИРБИС64 файл лежит в папке базы, запись ссылается на него полем 951. По документации ИРБИС64, переносить такие файлы вместе с записями «затруднительно», а при полном пути к файлу «возникнут проблемы при показе внешнего объекта через Web». Переезд сервера или смена АБИС рвут ссылки.

Как это устроено



Путь файла от съёмки до читателя и варианты хранения. Нажмите на схему, чтобы увеличить её.

Каталогизатор снимает издание, ИИ-каталогизация готовит запись и файлы. При передаче в АБИС запись получает ссылку на файл в формате каталога. Сам файл уходит в файловый архив через API — программный интерфейс для обмена между сервисами. Читатель открывает запись в каталоге или в личном кабинете, и каталог показывает файл по ссылке.

Где хранятся файлы — три варианта на схеме:

1. **В нашем файловом архиве библиотеки.** Адрес files.bibtech.ru. Сервис работает в Yandex Cloud, файлы лежат в Yandex Object Storage. Так устроено сейчас.
2. **В файловом архиве библиотеки в ЦОД библиотеки.** Хранилище стоит в центре обработки данных (ЦОД) библиотеки. ИИ-каталогизация отправляет файлы туда так же, через API.
3. **В хранилище библиотеки.** Файлы передаются туда из нашего хранилища по протоколу, согласованному с библиотекой.

Как файлы попадают в запись

Какие файлы идут в запись

1. На шаге «Выбор файлов» сотрудник отмечает галочкой «В запись» файлы, которые нужны в записи. Обложки отмечены заранее: сервис ставит галочку сам после обработки и когда сотрудник выбирает тип «Обложка» или «Готовая обложка». Ненужную обложку можно снять.
2. Если снять отметку со всех обложек, сервис подскажет «Обложка не включена в запись.» Ссылки на обложку в записи не будет.
3. У служебных снимков галочки нет, в запись они не идут. Это «Штрихкод», «Каталожная карточка» и разворот целиком. У разворота в запись идут его половины.
4. Сотрудник нажимает «Подтвердить проверку». После этого файлы записи не меняются: их нельзя переснять, удалить, снять с записи или сменить им тип. Если нужно поправить, то сначала нажмите «Отозвать проверку».

Как проверять файлы, рассказано в инструкции «Проверка и передача в АБИС».

Обложка или прочий файл

Вид файла сервис определяет по его типу. Отдельно выбирать вид не нужно.

- **Обложка** — «Обложка», «Готовая обложка». По ссылке открывается небольшой картинкой — для выдачи и карточки каталога.
- **Прочий файл** — «Титульная страница», «Содержание», «Задняя обложка», «Приложение», «Страница», файл без типа, PDF. Открывается крупно, чтобы читался текст.
- **Служебный** — «Штрихкод», «Каталожная карточка», «Разворот». В запись не идёт.

Обложкой считается только изображение. Обложка в PDF уходит в запись как прочий файл.

Строка ссылки в каталоге

Строку ссылки сервис строит так, как принято в каталоге библиотеки. В RUSMARC у некоторых каталогов свой вид строки — его настраиваем при подключении.

- **RUSMARC, по умолчанию.** Обложка — `856 41 $u <адрес> $q SRC`, прочий файл — `856 42 $u <адрес> $q SRC`.
- **RUSMARC, свой вид у одних каталогов** — `856 41 $q Cover $u <адрес>`, у обложки и прочего файла одинаково. Формат можно задать при интеграции.
- **MARC 21 (Koha)** — `856 40 $u <адрес> $3 <подпись>`, у обложки и прочего файла одинаково.
- **ИРБИС** — `951 ^I<адрес> ^T<подпись>`, у обложки и прочего файла одинаково.

Подпись в Koha и ИРБИС — название типа файла: «Обложка», «Титульная страница», «Содержание», «Задняя обложка», «Приложение». У файла без типа — «Страница». В строках RUSMARC подписи нет.

Адрес у обложки — `https://files.bibtech.ru/covers/<номер файла>.jpg`, у прочих изображений — `…/scans/<номер файла>.jpg`, у PDF — `…/scans/<номер файла>.pdf`. Номер файла — его код в ИИ-каталогизации. Ссылка ставится в запись сразу, даже если файл ещё не опубликован.

Повторная передача

- Свою строку сервис узнаёт по адресу и номеру файла и заменяет её на месте. Вид, раздел адреса и подпись могли смениться. Второй строки не появляется.
- Файл сняли с записи или заменили новым — его строка при передаче снимается.
- Новая обложка заменяет прежнюю ссылку на обложку, ссылки на страницы остаются. Так работают и «Добавление обложки», и «Фото обложки» при постановке на учёт.
- Задание без файлов «В запись» ссылок в записи не трогает.
- Строки 856 библиотеки со ссылками на другие хранилища не меняются, их порядок тоже.
- Строку, которую библиотека дополнила своими подполями, сервис не трогает: она считается строкой библиотеки.
- Ссылку, которую библиотека сама убрала из каталога, сервис не возвращает. Новая ссылка появится, только если сотрудник заменит файл.

Публикация файлов

Когда файл уходит в архив

Отмеченные файлы уходят в файловый архив, когда сотрудник:

- подтвердил проверку файлов или полей записи;
- передал запись в каталог;
- нажал «Принять».

Обычно файл публикуется за секунды. Передача записи публикации не ждёт: ссылка уже в записи, файл приходит следом.

Состояние файла

Состояние видно на плашке файла на шаге «Выбор файлов».

- **«Опубликован».** Файл в архиве, ссылка открывается. Делать ничего не нужно.
- **«В очереди».** Файл ещё не отправлен. Подождите, обычно это секунды.

- **«Не отправлен».** Отправка не удалась, причину см. в окне файла. Если файловый архив недоступен, повтор пойдёт сам, «Повторить» отправит сразу. Если причина — «Обратитесь в поддержку», повторять бесполезно: напишите в поддержку.

На шаге передачи то же самое написано под строкой 856:

- «Файл ещё не отправлен. Ссылка заработает после отправки.»
- «Файл не отправлен.» и причина, рядом «Повторить».
- «Публикация файлов не настроена. Ссылка будет в записи, файла по ней не будет.» Так бывает в режиме «Личное использование» и у библиотеки, для которой публикация не подключена.

Если файловый архив недоступен, отправка повторяется сама: через 1, 5 и 30 минут, через 2 и 12 часов, дальше раз в сутки. Кроме того, раз в сутки ИИ-каталогизация сверяет свои файлы со списком файлового архива и досылает недостающее.

Обложки и страницы — разного размера

Обложки и прочие файлы лежат в разных разделах архива: обложки — в `covers/`, остальные — в `scans/`. По ссылке из записи они открываются по-разному:

- обложка — шириной 200 точек, как миниатюра в выдаче и на карточке каталога;
- страница — шириной 1600 точек, чтобы читались оглавление и мелкий шрифт.

Если снимок меньше, он не увеличивается.

Замена файла и смена типа

- **Обложку пересняли или повторили её обработку после публикации.** Это новый файл с новой ссылкой. Прежний остаётся в архиве, и его ссылка в каталоге работает, пока запись не передана заново. Если прежний ещё не успел уйти, под плашками видно «Прежние версии файлов: 1 не отправлена» — он уйдёт сам.
- **Тип файла сменили после публикации.** Например, страницу отметили обложкой. Файл публикуется ещё раз, в разделе нового вида под тем же номером. Прежний адрес тоже работает. Строка в записи сменится при следующей передаче.
- **Задание передали в другую сеть.** Файл переезжает в хранилище новой сети, ссылка в каталоге остаётся прежней.

Как размещается файловый архив

Файловый архив — отдельный сервис на своём сервере, адрес files.bibtech.ru. От сервиса обработки он не зависит. Ссылки в каталогах работают, даже если

обработка остановлена или переехала. Записей каталога архив не хранит. Он хранит файлы и отдаёт их по адресу. Вместе с файлом ИИ-каталогизация передаёт наименование и номер задания и библиотеку-владельца. По ним сотрудник библиотеки находит свои файлы в архиве. Как это сделать, рассказано в инструкции «Просмотр файлов».

- **Запись по ключу сети.** ИИ-каталогизация отправляет файл через API с ключом сети библиотек. Ключ определяет, в чьё хранилище ляжет файл, — в чужое записать нельзя. Ключи не попадают ни в браузер сотрудника, ни в ответы сервисов.
- **Хранение в облаке.** Файлы лежат в объектном хранилище S3 — облачном хранилище файлов, у нас это Yandex Object Storage. У каждой сети библиотек свой бакет — отдельное закрытое хранилище в облаке. Из интернета бакет напрямую не открывается. Читает и пишет в него только архив.
- **Файл хранится без потерь**, таким, каким его прислали. Сжатие — только при показе.
- **Показ через image-прокси.** image-прокси — программа, которая на лету уменьшает картинку и переводит её в нужный формат. Ширина выбирается из заранее заданной сетки. У обложек — 100, 200, 400 и 800 точек, по умолчанию 200. У страниц — 800, 1200, 1600 и 2400 точек, по умолчанию 1600. У сети может быть своя ширина по умолчанию из той же сетки.
- **Другая ширина и формат.** Другую ширину каталог запрашивает в адресе: `https://files.bibtech.ru/w:400/covers/<номер файла>.jpg`. Формат задаёт расширение: `jpg`, `png`, `webp`, у обложек ещё `avif`.
- **Быстрый повторный показ.** Уменьшенная картинка сохраняется, поэтому второй раз архив не обращается к облаку.
- **Знак сети.** На картинках шириной от 400 точек в углу стоит знак сети — надпись или логотип на полупрозрачной подложке, как водяной знак. На миниатюрах 100 и 200 точек знака нет, на страницах он стоит на любой ширине. Знак ставится только при показе. Файл в хранилище и оригинал не меняются. По умолчанию ставим знак всегда. Библиотека может отказаться от вывода знака.
- **Оригинал** без обработки открывается только владельцам данных, с их адресов. Остальным он закрыт.
- **Сведения о снимке** — модель камеры, сведения о правах — при показе удаляются.
- **Адрес не меняется.** Адрес файла живёт столько же, сколько запись. Облако, бакет и программу показа можно сменить, ссылки в каталогах останутся прежними. Регистр букв в адресе не важен.
- **Перенос между сетями.** Задание передали в другую сеть — файл переезжает в бакет новой сети, адрес остаётся.

- **Само ничего не удаляется.** ИИ-каталогизация из файлового архива ничего не удаляет. Удаление файла или задания и снятая галочка копию в архиве не трогают. Удалить файл может только оператор архива отдельной командой.
- **Защита.** Архив работает по HTTPS. Число запросов с одного адреса ограничено, скорость скачивания больших файлов — тоже. Так архив не перегрузить.
- **Читатели** открывают файлы без входа.

На главной странице файлового архива files.bibtech.ru размещена справка.

Возможности файлового архива

Основные функции

- Хранит изображения JPEG, PNG, WebP, TIFF, а также PDF, аудио и видео — до 100 МБ на файл.
- Отдаёт по одному адресу картинку нужной ширины и формата.
- Ставит знак сети на крупные картинки.
- Держит файлы каждой сети в отдельном бакете — учёт и хранение по сети.
- Ведёт журнал операций: какой файл, когда и каким ключом записан, удалён или перенесён.
- Считает по сетям число файлов и объём хранения.
- Сверяет свой список файлов с бакетом и находит пропавшие. Пропавший файл ИИ-каталогизация отправляет заново, адрес остаётся прежним.
- Переносит файлы между сетями без смены адреса.
- Показывает сотрудникам библиотек их пакеты и файлы: реестр, галерея, статистика. Вход — кнопка «Администрирование» на главной странице архива. Как работать, рассказано в инструкции «Просмотр файлов».
- Предел объёма хранения для сети - 10 Тб, если нужно более сообщите.

Расширяя по запросу от библиотек

По запросу от библиотек возможно:

- **Свой файловый архив у сети или каталога.** - библиотека может развернуть сервис на собственных вычислительных мощностях.
- **Новая сеть библиотек** — свой бакет, ключ и настройка. ИИ-каталогизация начинает отправлять туда файлы её библиотек, а сверка досылает файлы записей, переданных раньше.
- **Свой показ у сети** — ширина по умолчанию, знак сети (надпись или логотип), его угол и ширина, с которой он появляется.
- **Сетка ширин и форматы** — настройка раздела архива. Ссылки в каталогах при этом не меняются.

- **Своё облако сети** — свой аккаунт или свой поставщик хранилища S3. Переезд — копирование файлов и смена настройки, адреса остаются.
- **Свой домен сети**, например `files.<библиотека>.ru`. Адреса файлов внутри не меняются.
- **Раздача через облачного поставщика** — балансировщик или сеть доставки с его сертификатом HTTPS вместо нашего.
- **Запись своими системами сети** — ключ сети можно выдать системам самой сети.

Частые вопросы

Ссылка в записи есть, а файл не открывается

Посмотрите состояние файла на шаге «Выбор файлов». «В очереди» — файл ещё не отправлен, подождите. «Не отправлен» — причина в окне файла, нажмите «Повторить». «Публикация файлов не настроена» — у вашей библиотеки публикация не подключена: ссылка в записи есть, а файла по ней нет. В последнем случае напишите в поддержку.

Как убрать из записи страницу и оставить только обложку?

Если проверка файлов подтверждена, нажмите «Отозвать проверку». Снимите галочку «В запись» у страницы и передайте запись заново. Ссылка на страницу уйдёт из записи. Файл в файловом архиве останется: ИИ-каталогизация его не удаляет.

Почему в старой записи сервис показывает изменение строки 856?

Запись передали до того, как обложки и страницы разделились. Сервис показывает, как строка выглядит теперь: новый вид, у страниц — адрес в разделе `scans/`. Передавать запись заново не обязательно, решите сами. До повторной передачи каталог показывает страницы по прежнему адресу, небольшой картинкой.

Что будет со ссылками, если библиотека перейдёт в другую сеть?

Файлы переедут в хранилище новой сети. Адреса останутся прежними, записи править не нужно.

Удалится ли файл из архива, если удалить задание?

Нет. Удаление задания или файла и снятая галочка копию в файловом архиве не трогают. Ссылка, которая уже ушла в каталог, продолжает работать.

Можно ли получить оригинал снимка?

Оригинал без обработки открывается только владельцам данных, с их адресов. По обычной ссылке доступна картинка: самая крупная — 800 точек у обложки и 2400 у страницы.