# ОРГАНИЗАЦИЯ, ФОРМИРОВАНИЕ И СОХРАННОСТЬ ФОНДОВ

УДК 002.1 – 028.27 + 025.13:004 + 004.352:025.2
https://doi.org/10.33186/1027-3689-2026-9-43-69

## Селективная потоковая оцифровка крупных библиотечных фондов

**И. В. Морозов[1], Н. А. Михальченкова[2]**

*[1, 2]ГПНТБ России, Москва, Российская Федерация*

*[1]morozoviv@gpntb.ru, https://orcid.org/0009-0004-3358-5372*
*[2]MihalchenkovaNA@gpntb.ru*

***Аннотация.*** При массовой оцифровке библиотечных фондов возникает известная проблема – планетарное сканирование оставляет книгу неповрежденной, но остается медленным и неэффективным при оцифровке фондов, насчитывающих миллионы единиц хранения. В статье рассматривается другой подход – селективная потоковая оцифровка, при которой у специально отобранного дублетного экземпляра отделяется корешковая часть, а полученные листы обрабатываются на скоростном потоковом сканере. Авторы сравнивают этот подход с традиционным постраничным сканированием и показывают, в каких случаях разборка экземпляра является оправданной, а в каких – недопустимой.

Расчетное моделирование показывает, что для фонда в 1 млн изданий этап цифрового захвата сокращается примерно с 21 года до одного года – семи месяцев в зависимости от класса оборудования, то есть в 21–35 раз.

Отдельное место в работе занимают вопросы экземплярного учета и координации процессов оцифровки. Если несколько библиотек независимо друг от друга оцифровывают одни и те же издания, возникают дублирующие копии. Национальная электронная библиотека (НЭБ), действующая как федеральная государственная информационная система, обеспечивает хранение цифровых копий и ведение объединенного электронного каталога фондов российских библиотек. Однако учет в ней, как и в других сводных каталогах, ведется на уровне библиографической записи, тогда как селективная потоковая оцифровка требует сведений о состоянии физического экземпляра и его дублетности. Поэтому авторы предлагают создавать не новую систему экземплярного учета, а подсистему экземплярного учета и координации оцифровки в составе НЭБ, в задачи которой будут входить выявление дублетных изданий, сквозная

идентификация через универсальные классификаторы и идентификаторы, учет уже оцифрованных экземпляров с указанием мест хранения и формирование на этой основе массива изданий для промышленного сканирования.

Отдельно анализируется зарубежный опыт – от проектов China Academic Digital Associative Library и Digital Library of India до коллективного судебного иска группы американских авторов к компании Anthropic, в котором суд разграничил перевод книг в цифровую форму методом потоковой оцифровки и их использование при обучении языковых моделей искусственного интеллекта. С учетом вступающего в силу Федерального закона от 26 июля 2026 г. № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации» авторы показывают, как масштабная потоковая оцифровка может стать источником для обучения отечественных моделей искусственного интеллекта без ущерба интересам правообладателей.

Статья подготовлена в рамках государственного задания, утвержденного Минобрнауки России для ГПНТБ России на 2026 год по выполнению НИОКТР Рег. № 126011215424-5 по теме «Открытый доступ в научно-информационной деятельности и разработка прототипа многофункциональной системы открытого архива хранения и аналитической обработки научно-технической информации с учетом механизмов открытых лицензий и элементов искусственного интеллекта» (FNEG-2025-0004).

***Ключевые слова:*** оцифровка библиотечных фондов, деструктивная оцифровка, селективная оцифровка, потоковое сканирование, дублетные экземпляры, сводные каталоги, Национальная электронная библиотека, подсистема экземплярного учета, ГСНТИ, научно-техническая информация, большие языковые модели, обучение искусственного интеллекта, Федеральный закон от 26 июля 2026 г. № 243-ФЗ, авторское право

***Для цитирования:*** Морозов И. В., Михальченкова Н. А. Селективная потоковая оцифровка крупных библиотечных фондов // Научные и технические библиотеки. 2026. № 9. С. 43–69. https://doi.org/10.33186/1027-3689-2026-9-43-69

# COLLECTION MANAGEMENT, DEVELOPMENT AND PRESERVATION

UDC 002.1 – 028.27 + 025.13:004 + 004.352:025.2
https://doi.org/10.33186/1027-3689-2026-9-43-69

## Selective high-throughput digitization of large library collections

**Ivan V. Morozov[1] and Natalia A. Mikhalchenkova[2]**

*[1, 2]Russian National Public Library for Science and Technology, Moscow, Russian Federation*

*[1]morozoviv@gpntb.ru, https://orcid.org/0009-0004-3358-5372*
*[2]MihalchenkovaNA@gpntb.ru*

***Abstract.*** Large-scale digitization of library collections presents a dilemma: planetary scanning leaves the book undamaged, yet it remains slow and inefficient when millions of items must be processed. The authors examine an alternative approach — selective high-throughput digitization, in which a duplicate copy is selected, its spine is removed, and the resulting sheets are processed on a high-performance sheet-fed scanner. The authors compare this approach with traditional page-by-page scanning and show under what conditions disbinding a copy is justified and when it is inadmissible.

Modelling shows that for a collection of one million volumes the image-capture stage can be reduced from approximately 21 years to between one year and seven months, depending on scanner class — a 21- to 35-fold gain.

The study also focuses on item-level accounting and the coordination of digitization. When several libraries independently digitize the same titles, duplicate digital copies are produced. The National Electronic Library (NEL), operating as a federal state information system, already provides for the storage of digital copies and maintains the union electronic catalogue of Russian library collections. However, like other union catalogues, it offers record-level bibliographic control, whereas selective high-throughput digitization requires data on the physical condition of an individual item and its duplicate status. The authors therefore propose not a new national system, but an item-level accounting and digitization coordination subsystem within the existing NEL infrastructure. Its tasks would include identifying duplicate editions, providing end-to-end identification through universal classifiers and identifiers, recording already digitized copies together with their location data, and, on this basis, compiling the array of publications for industrial-scale scanning.

The authors also review international experience, from the China Academic Digital Associative Library and Digital Library of India projects to the class action brought by a group of American authors against Anthropic (Bartz et al. v. Anthropic PBC), in which the court distinguished between converting books into digital form by high-throughput digitization and using them to train artificial intelligence language models. In the light of Federal Law No. 243-FZ of 26 July 2026 "On Supporting the Development of Artificial Intelligence Technologies in the Russian Federation", the authors show how large-scale high-throughput digitization may become a source of data for training domestic AI models without prejudice to the interests of rightsholders.

The paper is prepared under the Government Order approved by the Ministry of Science and Higher Education of the Russian Federation to RNPLS&T for 2026, R&TD theme Reg. No. 126011215424-5 "Open access in scientific information activities and development of a prototype of a multifunctional open archive system for storage and analytical processing of scientific and technical information taking into account open licensing mechanisms and elements of artificial intelligence" (FNEG-2025-0004).

***Keywords:*** library collection digitization, destructive digitization, selective digitization, high-throughput scanning, disbinding, duplicate copies, union catalogues, National Electronic Library, item-level accounting subsystem, State System of Scientific and Technical Information (GSNTI), scientific and technical information, large language models, artificial intelligence training, Federal Law No. 243-FZ, copyright

***Cite:*** Morozov I. V., Mikhalchenkova N. A. Selective high-throughput digitization of large library collections // Scientific and technical libraries. 2026. No. 9, pp. 43–69. https://doi.org/10.33186/1027-3689-2026-9-43-69

## Введение

Сегодня для библиотек одной из приоритетных задач является оцифровка книжных фондов. Современные технологии – интеллектуальный и семантический поиск, интеллектуальный анализ, использование средств ИИ – расширяют возможности оцифровки. Наиболее распространенной не только в нашей стране, но и за рубежом является стандартная технология сканирования с использованием планетарных сканеров или цифровых фотокамер. Например, FADGI – одна из ведущих международных инициатив в области стандартизации оцифровки, относит планетарные книжные сканеры к рекомендуемым технологиям

сканирования для переплетных изданий [**1**]. Но в тех же рекомендациях есть важная оговорка: если книжный том уже технологически разобран и у него удалена корешковая часть, он может быть оцифрован на скоростном потоковом сканере.

Именно это является темой данной статьи. Обязательно ли использовать технологию планетарного сканирования, которая не предусматривает повреждение изданий, для всего фонда? Если по фондам разных библиотек «разбросано» одно и то же издание из тиража в сотни или тысячи экземпляров, является ли оправданной технологическая разборка дуплетного экземпляра, то есть повреждение одного из дуплетных экземпляров ради значительного увеличения скорости оцифровки всего фонда? С этими вопросами сталкиваются сейчас библиотеки России и ГПНТБ России в частности.

**Цель исследования** – предложить организационно-технологическую модель селективной потоковой оцифровки крупных библиотечных фондов, оценить ее преимущества на примере крупного библиотечного фонда и определить, при каких условиях технологическая разборка экземпляра оправдана как способ значительно увеличить производительность. Вторая задача исследования – показать, что промышленное применение этой технологии требует координации и экземплярного учета, которого действующая национальная инфраструктура пока не имеет, и мы предлагаем ее планирование как подсистему домена «Наука и инновации».

**Научная новизна** заключается в том, что высокоскоростная потоковая оцифровка рассматривается нами не как отдельная технология, а как часть более широкого подхода, который состоит:

а) из экземплярного отбора документов;

б) разделения технологий сканирования;

в) координации процесса оцифровки на национальном уровне на основе экземплярного учета;

г) взаимосвязи авторского права и процесса потоковой оцифровки, в том числе для обучения моделей искусственного интеллекта массивом оцифрованных данных.

## Материалы и методы исследования

Работа выполнена как аналитическое и расчетное исследование, а не как практический эксперимент. Задача авторов – обосновать организационно-технологическую модель и оценить потенциал исследуемой технологии до того, как она будет массово внедрена на практике. В основе работы лежат четыре метода: сравнительный анализ способов оцифровки, изучение практических примеров зарубежного опыта, моделирование и расчетная оценка производительности.

База методических материалов включает несколько источников. Во-первых, это технические рекомендации FADGI как основной международный ориентир по стандартам оцифровки [**1**]. Во-вторых, материалы OCLC Research[1] по высокопроизводительной оцифровке специальных и архивных коллекций [**2**]. В-третьих, публикации по крупным зарубежным проектам оцифровки: China-US Million Book Digital Library Project [**3**], CADAL [**4**] и Digital Library of India [**5**], а также материалы Wellcome Library [**6**]. Отдельную группу составляют судебные материалы по делу Bartz et al. v. Anthropic PBC [**7**], где на примере зарубежного опыта дана правовая оценка потоковому сканированию и использованию полученных цифровых копий. Для расчетной части исследования проанализированы технические характеристики современных потоковых сканеров.

## Моделирование производительности

Для моделирования производительности рассмотрим планетарный способ сканирования книги объемом 250 стр. (125 разворотов). Согласно паспортным характеристикам планетарных сканирующих комплексов [**8**], номинальная производительность оборудования этого класса достигает 1 тыс. стр. в час (время срабатывания камеры составляет доли секунды, а технический цикл съемки одного разворота – 3–4 секунды). Однако паспортная скорость отражает работу камеры и не учитывает ручные операции: перелистывание, выравнивание страниц на колыбели сканера, контроль геометрии раскрытия, визуальную проверку кадра. По данным центра сканирования ГПНТБ России, фактическая выработка

---

[1] OCLC Research – исследовательское подразделение международного некоммерческого библиотечного консорциума OCLC, разрабатывающее технологические стандарты и модели межбиблиотечного взаимодействия.

при соблюдении требований к качеству составляет четыре разворота (8 стр.) в минуту – около 480 стр. в час. При такой производительности оцифровка тома объемом 250 стр. занимает порядка 30 минут, причем округление выполнено в меньшую сторону. Отметим, что общеотраслевого норматива не существует, действующие нормы времени прямо указывают, что выработку при сканировании следует определять для конкретного рабочего места на основе хронометража, поскольку она зависит от мощности оборудования и сложности материала [**9**].

Представим, что библиотека располагает фондом в 1 млн экз. и десятью планетарными сканерами. При восьмичасовом рабочем дне и 24 рабочих днях в месяц на полную оцифровку такого фонда силами одного учреждения уйдет примерно 21 год! И это при условии непрерывной работы без поломок оборудования и отпусков сотрудников. Кроме колоссальных временных затрат это и серьезная нагрузка на бюджет. Даже поверхностный финансовый анализ, опирающийся на примерные рыночные данные оплаты труда операторов сканирования [**10**], позволяет определить, что совокупный фонд оплаты труда этих десяти сотрудников за 21 год составит порядка 234 млн руб. И это «чистыми», без учета обязательных страховых взносов, закупки оборудования, расходов на электричество и серверные мощности для хранения терабайтов данных.

Сразу оговоримся, что авторы данной статьи не претендуют на академическую точность или построение строгой экономической модели. Наша цель – представить условный практический расчет, который демонстрирует масштаб технологических затрат.

Теперь сравним производительность планетарных сканеров с производительностью современных двусторонних потоковых аппаратов. Потоковый сканер устроен иначе. Он подает листы автоматически, после чего сканирует их сразу с обеих сторон. Кроме того, современные модели оснащены встроенными механизмами автоматического оптического распознавания символов (OCR), что позволяет получать распознанный текст в процессе сканирования. Производительность здесь определяется не числом ручных операций на страницу, а скоростью, с которой устройство подачи прогоняет листы через сканер.

Возьмем те же исходные данные – фонд в количестве 1 млн экз., а десять планетарных сканеров заменим таким же количеством современных двусторонних потоковых аппаратов. Расчет показывает оше-

ломляющий разрыв в эффективности. Оборудование класса сканирования 90 листов в минуту, например, Epson DS-32000 [**11**], способно обработать одну книгу объемом 250 стр. (125 отдельных листов, полученных после технологической разборки тома) всего за 1,4 минуты. При тех же вводных данных (восьмичасовой рабочий день, 24 рабочих дня в месяц) на оцифровку миллионного фонда потребуется ровно год вместо ранее рассчитанного 21 года. Использование же более производительных систем класса 150 листов в минуту, например, PFU/Ricoh fi-8950 [**12**], сокращает время с одного года до 7,2 месяца. Таким образом, производительность потоковых систем[2] превосходит ручное планетарное сканирование в 21–35 раз.

Следует добавить, что разборка книги, подготовка листов, обработка обложек и вкладок, проверка последовательности и контроль качества требуют дополнительного времени и не входят в расчет. Поэтому диапазон 21–35 раз стоит воспринимать именно как моделирование, а не как готовый ответ. Реальный коэффициент повышения производительности можно будет оценить только в ходе экспериментальной проверки. Но такая разница в эффективности принципиально важна даже на стадии моделирования. В масштабах оцифровки многомиллионных фондов увеличение скорости даже одного этапа технологической цепочки способно кардинально изменить сроки и экономику всего процесса. В масштабах страны такое повышение эффективности можно считать внутриотраслевой промышленной революцией.

**От постраничного сканирования к потоковой модели**

Чтобы издание можно было отправить в потоковый сканер, книжный блок нужно превратить в упорядоченный набор отдельных листов. На рис. 1а и 1б представлен один из вариантов технологической разборки переплетенного издания на типографском резательном оборудовании ГПНТБ России для последующего потокового сканирования. Корешковая часть книжного блока отделяется с использованием гидравлического резательного оборудования, обложки удаляются, после чего издание превращается в однородную стопку отдельных листов, готовую для автоматической подачи в потоковый сканер. При этом,

---

[2] Указанные модели потоковых сканеров приводятся в качестве примера и не являются обязательной рекомендацией.

если обложка или переплет издания представляют самостоятельную библиографическую или художественную ценность, они могут быть оцифрованы отдельно на традиционном планетарном сканере.

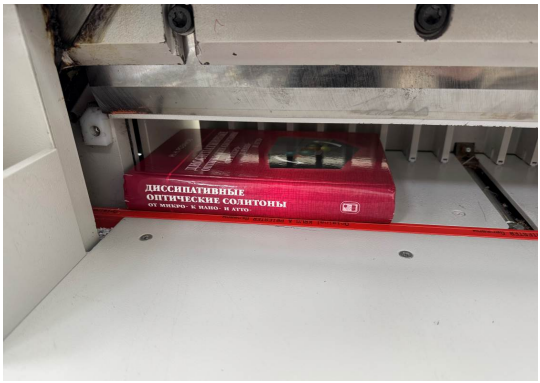



**Рис. 1а. Технологическая подготовка переплетенного издания к потоковой оцифровке: размещение книжного блока на типографском резательном оборудовании**




**Рис. 1б. Технологическая подготовка переплетенного издания к потоковой оцифровке: отделение корешковой части на типографском резательном оборудовании**

После отделения корешковой части и удаления обложек книга превращается в обычную стопку несвязанных страниц (рис. 1в). Это позволяет начать высокоскоростную потоковую оцифровку.

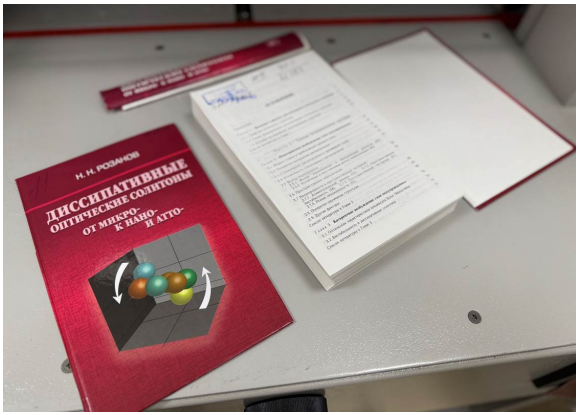



**Рис. 1в. Технологическая подготовка переплетенного издания к потоковой оцифровке: технологически разобранный книжный блок**

Общая последовательность операций показана на рис. 2.




**Рис. 2. Последовательность технологических операций при селективной потоковой оцифровке переплетенного издания**

Как показано на рис. 2, предлагаемая модель включает четыре основные стадии:

а) отбор исходного экземпляра,

б) технологическую обработку книжного блока на резательном оборудовании с целью отделения корешковой части,

в) формирование однородной стопки отдельных листов,

г) потоковое двустороннее сканирование.

При технологической разборке физическая конструкция экземпляра нарушается, поэтому в зарубежной практике этот подход часто называют деструктивной оцифровкой (англ. destructive digitization) [**13**]. По мнению авторов, этот термин является не совсем корректным. Если один экземпляр из многочисленных дублетов в библиотечных фондах не имеет самостоятельной исторической или культурной ценности, а его разрушение обеспечивает колоссальный прирост скорости и значительное сокращение временных, организационных и финансовых затрат, то это является не только оправданным, но и необходимым. К тому же информация из издания не уничтожается, а меняется лишь форма ее представления. В цифровом виде данный документ принесет несоизмеримо больше пользы, в то время как бумажный дублет зачастую выполняет лишь роль архивного хранения, подвергаясь рискам естественного старения, физического разрушения из-за сырости или плесени, и в конечном итоге безвозвратному списанию и утилизации. Поэтому в настоящей работе авторы предпочитают термин «селективная (выборочная) оцифровка». Такой подход к способам оцифровки по четким критериям (состояние документа, его ценность, наличие других экземпляров) зафиксирован в руководстве IFLA/UNESCO по проектам оцифровки [**14**].

Практически это означает разделение фонда на три технологических маршрута. Первый – недеструктивная оцифровка, которой подлежат книжные памятники, рукописи, редкие и особо ценные издания, единственные экземпляры и хрупкие материалы, обрабатывающиеся только планетарным способом. Второй – селективная потоковая оцифровка для массовых тиражных дублетов, не имеющих самостоятельной ценности и пригодных для автоматической листовой подачи. Третий – гибридный вариант, в котором части книги, представляющие артефактную ценность (например, обложка) сканируются планетарно, а все

остальные – потоковым сканированием. Отнесение экземпляра к конкретному маршруту определяется по результатам экспертизы.

**Зарубежный опыт массовой оцифровки**

Уникальным примером массовой оцифровки является китайско-американский проект China-US Million Book Project [**3**], стартовавший в 2001 г. и со временем переросший в крупнейшую национальную цифровую инфраструктуру Китая CADAL (China Academic Digital Associative Library) [**4**]. В рамках этой инфраструктуры, объединившей более 70 университетов, процессы оцифровки были разделены: тиражная академическая литература XX в. обрабатывалась скоростным потоковым сканированием, а редкие фонды и памятники письменности проходили процедуру строгого отбора и бережного сканирования. По данным Zhejiang University Libraries, к 2012 г. в CADAL было оцифровано порядка 2,5 млн объектов. Для решения проблемы дублирования оцифровки были распределены зоны ответственности между участниками: за каждым крупным университетским центром закреплялись конкретные тематические или хронологические массивы литературы. Такой подход позволил исключить повторную оцифровку идентичных изданий.

С организационной точки зрения интересен опыт Digital Library of India – масштабного государственного проекта Правительства Индии по оцифровке и сохранению культурного, исторического и научного наследия, в рамках которого активно применялись технологии потокового сканирования [**15**]. В рамках этого проекта на базе регионального центра оцифровки в Хайдарабаде [**16**] была развернута масштабная инфраструктура, включающая 50 специализированных сканирующих комплексов. Высокопроизводительное оборудование применялось только к обычным печатным книгам, рукописи на пальмовых листьях и другие хрупкие уникальные материалы фотографировались с использованием цифровых фотокамер. Проблему дублирования оцифровки решали просто – библиотекарь проверял данные каждого издания в центре координации на наличие дубликата в общей базе, после чего допускалось сканирование [Там же]. Авторы индийского проекта рассматривали дублирование при оцифровке не как побочный эффект, а как

самостоятельную проблему, напрямую снижающую производительность всей системы.

По-своему интересен и европейский опыт. В Wellcome Library [**6**] подготовка изданий к оцифровке проходила двумя способами: разборка тома контролируемым демонтажем переплета (disbinding bound collections)[3] либо полное исключение издания из процесса потокового сканирования в случае, если оно обладало уникальными элементами оформления или обложка имела самостоятельную артефактную ценность. Схожие подходы описаны в Ватиканской апостольской библиотеке, Национальном архиве Великобритании и ряде других европейских учреждений. Их объединяет строгое разделение информационной и материальной ценности экземпляра. Выбор технологии оцифровки зависел от конкретной книги, что исключало применение однотипных технических решений в ущерб сохранению ценности [**17**].

Отдельного рассмотрения заслуживает американская практика, но уже в другом контексте, поэтому к ней мы вернемся ниже.

Итак, если обобщить опыт Китая, Индии и Европы, можно сделать два важных вывода. Во-первых, промышленная оцифровка требует заранее классифицировать объекты по физическим и экземплярным характеристикам. Универсальной технологии, одинаково подходящей для массового тиражного издания, редкой книги, архивного документа и уникальной рукописи, не существует. Во-вторых, при переходе к массовой оцифровке важна не только скорость работы оборудования, но и координация между библиотеками и центрами научно-технической информации. Зарубежный опыт показывает, что без единого механизма проверки уже сделанной работы одни и те же издания оцифровываются повторно, а это приводит к потере ресурсов.

**От проблемы пересечения фондов к национальной координации**

Согласно исследованию, проведенному экспертами OCLC Research в 2013 г., более 60% печатных фондов ведущих исследовательских библиотек должны иметь дублирующую копию в цифровых хранили-

---

[3] Disbinding bound collections (англ. демонтаж переплетенных коллекций) – бережное ручное разъединение книжного блока (распарывание нитей, размягчение клея), предполагающее последующее восстановление и переплетение книги.

щах [**18**]. Однако, по более поздним и осторожным оценкам, медианное пересечение фондов оказалось скромнее – порядка 30–35%, максимум 50% для отдельных библиотек. В нашей стране ситуация выглядит немного иначе. Проблема заключается в предельно низком охвате оцифровки как таковой. Например, в Российской национальной библиотеке (РНБ), фонд которой превышает 40 млн единиц хранения, оцифровано свыше 650 тыс. документов [**19**], то есть около 1,5–2% фонда. Национальная электронная библиотека (НЭБ), агрегирующая оцифрованные фонды 60 российских библиотек, включая РГБ, РНБ и ГПНТБ России, насчитывает порядка 5,8 млн оцифрованных документов [**20**]. Для России с ее огромными фондами (только у РГБ – свыше 47 млн единиц хранения [**21**]) эта доля остается довольно скромной. При столь низком общем охвате оцифровки цена каждой продублированной, повторно оцифрованной книги оказывается непропорционально высокой.

Необходимо оговориться: утверждать, что в России отсутствует национальная инфраструктура оцифровки библиотечных фондов, было бы неверно. Такая инфраструктура создана и закреплена законодательно статьей 18[1] Федерального закона «О библиотечном деле», где НЭБ определена как федеральная государственная информационная система, ее оператором назначена РГБ, а участниками являются государственные и муниципальные библиотеки, библиотеки образовательных и научных учреждений [**22**]. Более того, на оператора НЭБ прямо возложено ведение объединенного электронного каталога фондов российских библиотек. Параллельно центром ЛИБНЕТ ведется Сводный каталог библиотек России (СКБР) – крупнейший централизованный сводный электронный каталог страны, охватывающий фонды более 500 библиотек [**23**]. Библиотеки и научные центры регулярно передают в НЭБ цифровые копии, создаваемые в рамках государственных заданий, и делают это преимущественно средствами планетарного сканирования.

Таким образом, вопрос не в отсутствии национальной цифровой инфраструктуры, а в полноте информации об издании, которую она обеспечивает. СКБР и объединенный каталог НЭБ описывают фонды на уровне библиографической записи, они отвечают на первую часть во-

проса, какое издание существует, в каких организациях оно хранится и есть ли у него электронная копия. Потоковая оцифровка требует ответа на вторую часть – сколько физических экземпляров конкретного издания хранится в разных библиотеках, в каком они состоянии и какое из них допустимо направить на технологическую разборку. Эти сведения остаются во внутренних информационных системах отдельных библиотек и на национальном уровне не обрабатываются.

Если же говорить о фондах, которые еще не оцифрованы, то в российском библиотековедении для количественной оценки их пересечения предложен формальный аппарат на основе теории множеств. Фонд каждой библиотеки рассматривается как множество документов, а степень их однородности оценивается через попарное пересечение этих множеств [**24**]. Сам факт появления такого исследования в профильной литературе говорит о том, что проблема дублирования фондов осознается отечественным библиотечным сообществом. При этом задача координации не нова для отечественной практики, ее решение состоит из нескольких этапов. На первом этапе постановлением Правительства РФ от 24 июля 1997 г. № 950 была создана Государственная система научно-технической информации (ГСНТИ), объединявшая ГПНТБ России, ВИНИТИ, отраслевые и региональные органы НТИ по всей стране [**25**]. Но ГСНТИ строилась как организационно-правовое объединение учреждений, а не как техническая инфраструктура. Она регулировала взаимодействие организаций, а не информационные системы. Вторым этапом стало создание НЭБ – уже полноценной федеральной информационной системы с общим фондом цифровых копий и объединенным каталогом. Третий этап – экземплярный учет и координация оцифровки – пока не пройден, поскольку ни на первом, ни на втором этапе такая задача не ставилась. ГСНТИ работала на уровне организаций, НЭБ работает на уровне изданий и цифровых объектов.

После проделанной работы было бы вполне логично создать подсистему экземплярного учета и координации оцифровки по опыту НЭБ как подсистемы домена «Наука и инновации», интегрированного с другими подсистемами сводной каталогизации держателей фондов, в том числе органов Государственной системы научно-технической информации. Там уже есть правовая основа, инфраструктура хранения циф-

ровых копий, налажено взаимодействие с библиотеками-участниками. Создание такой подсистемы приобретает дополнительный вес в связи с принятием Федерального закона от 26 июля 2026 г. № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации» [**26**]. Пункт 5 части 2 статьи 5 этого закона прямо предусматривает установление Правительством РФ порядка предоставления доступа к информации (составам данных), содержащейся в том числе в информационных системах федеральных государственных учреждений, в целях обучения суверенных и (или) национальных больших фундаментальных моделей искусственного интеллекта.

Возникает прямая связь с темой настоящей статьи – оцифрованные библиотечные фонды, прошедшие распознавание текста, являются одним из источников машиночитаемых массивов данных. Российская практика уже движется в этом направлении – в июне 2026 г. РГБ объявила о запуске ИИ-платформы, в рамках которой фонд библиотеки становится доступен не только в виде полных текстов, но и в формате структурированных датасетов, пригодных для обучения языковых моделей [**27**]. В сегменте научно-технической информации у такой задачи есть естественная экспериментальная площадка в лице ГПНТБ России, где функционирует Российский сводный каталог по научно-технической литературе [**28**] и развивается Интегрированный сводный каталог научно-технической информации [**29**]. Как и СКБР, они работают на уровне библиографической записи, поэтому экземплярный учет и координацию оцифровки можно отработать до масштабирования на национальном уровне. Дополнительный вес этому придает федеральный проект по развитию научно-технических библиотек на 2025–2030 гг., реализуемый во исполнение Поручения Президента Российской Федерации от 30 марта 2024 г. № Пр-616 [**30**], среди направлений которого – создание реестра НТБ и центров научных знаний.

**Правовые и организационные ограничения**

То, что разборку экземпляра технологически возможно осуществить, еще не значит, что ее автоматически можно применять к документам библиотечного фонда. Действующий Порядок учета документов, входящих в состав библиотечного фонда, утвержденный приказом

Минкультуры России от 8 октября 2012 г. № 1077, регулирует, в частности, процедуры выбытия документов [**31**]. Он продолжает применяться в библиотечной практике в действующей редакции. Поэтому специальное технологическое использование экземпляра, состоящего на учете в основном фонде, требует отдельной юридической и организационной проработки. Наличие множества аналогичных экземпляров в библиотеках само по себе не дает оснований произвольно менять физическую структуру документа.

Не менее важны и ограничения авторского права. Статья 1275 ГК РФ действительно предусматривает специальные случаи, когда библиотеки могут создавать электронные копии без согласия правообладателя, но эти случаи ограничены целями и категориями произведений, прямо установленными законом [**32**]. И важно не путать одно с другим, потому что право создать копию – это не то же самое, что право неограниченно ее распространять или выкладывать в открытый доступ [**33**].

Существенно меняет правовой режим использования произведений при обучении систем искусственного интеллекта упомянутый выше Федеральный закон № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации» [**26**]. Часть 2 статьи 10 этого закона устанавливает, что обращение к информации, содержащейся в объектах авторского и смежного права, включая компьютерную обработку, извлечение, сравнение, классификацию и анализ закономерностей, а также воспроизведение посредством краткосрочной записи в память ЭВМ не является нарушением авторских и смежных прав, если это делается исключительно для обучения суверенных и (или) национальных больших фундаментальных моделей искусственного интеллекта и при условии, что разработчик правомерно получил экземпляр произведения либо оно правомерно общедоступно без технических ограничений[4].

Один из самых показательных примеров такого рода – проект американской компании Anthropic[5], известный под названием Project

---

[4] Данная норма вступает в силу с 1 марта 2027 г.

[5] Anthropic (Anthropic PBC) – американская компания по разработке больших языковых моделей искусственного интеллекта.

Panama [**34**]. Суть этого проекта заключалась в массовой скупке тиражных книг через крупные торговые сети. Физические экземпляры доставлялись на специализированные промышленные склады, где с помощью гидравлического резательного оборудования у книг полностью отделялась корешковая часть. Полученный массив листов пропускался через высокоскоростные потоковые сканеры, после чего бумажные оригиналы утилизировались, а оцифрованная распознанная (машиночитаемая) информация использовалась для обучения больших языковых моделей искусственного интеллекта.

Раскрытие информации о данном проекте привело к масштабному коллективному иску со стороны профессионального сообщества писателей (дело Bartz et al. v. Anthropic PBC[6], рассмотренное в Федеральном суде США) [**7**]. Суд отдельно рассмотрел использование цифровых копий произведений для обучения больших языковых моделей и оцифровку законно приобретенных печатных экземпляров для внутренней библиотеки Anthropic, и признал оба сценария правомерными в рамках доктрины добросовестного использования (fair use), однако разграничил правовые основания для каждого из них. Использование произведений для обучения языковых моделей было квалифицировано как процесс глубокой функциональной модификации, изменяющий целевое назначение книги. В свою очередь оцифровка легально выкупленного печатного экземпляра была оценена судом как конвертация формата данных в границах одного собственника, не повлекшая за собой создания дополнительных копий для внешнего коммерческого распространения. Важная оговорка – это решение не распространялось на библиотеку, сформированную из цифровых копий, полученных из неправомерных источников.

Данный пример основан на нормах американского авторского права, поэтому не может автоматически переноситься на российское законодательство, но для библиотечной практики это показательный пример, который демонстрирует, что цифровая копия создается не только для чтения человеком или удаленного доступа, но и для машин-

---

[6] Bartz et al. v. Anthropic PBC (от лат. et alii «и другие», v./versus «против») – коллективный судебный иск группы американских авторов (возглавляемой писателем Андреа Барцем) к компании Anthropic.

ной обработки, в том числе для обучения моделей искусственного интеллекта [**35**].

Авторы сочли важным отдельно остановиться на этом примере по двум причинам. Во-первых, в основе описываемого выше примера лежит именно потоковая технология оцифровки. Во-вторых, в России принят Федеральный закон от 26 июля 2026 г. № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации» [**26**], на котором авторы уже останавливались в данном исследовании и который напрямую затрагивает вопрос использования оцифрованных библиотечных фондов для обучения отечественных моделей искусственного интеллекта. Применительно к библиотечной оцифровке необходимо разграничивать минимум три самостоятельных вопроса: допустимость технологического использования конкретного экземпляра, правомерность создания и хранения его электронной копии, и правовой режим дальнейшего использования этой информации, в том числе для обучения систем искусственного интеллекта.

**Заключение**

Проведенное исследование обосновывает перспективность потоковой оцифровки с технологической разборкой специально отобранного экземпляра как способа обработки массовых тиражных изданий в составе крупных библиотечных фондов. Это подтверждают и зарубежные методические материалы с проектами, реализованными на практике.

Расчетное моделирование при переходе от планетарного сканирования к двусторонней потоковой обработке показывает сокращение времени оцифровки миллионного библиотечного фонда с 21 до одного года и меньше (7,2 месяца), в зависимости от класса оборудования. Это соответствует приращению скорости примерно в 21–35 раз. Эти результаты можно использовать как предварительную оценку, но не как доказанную характеристику полного производственного цикла.

Одним из главных условий являются экземплярный учет и координация оцифровки, позволяющие понять, сколько экземпляров того или иного издания существует в фондах разных библиотек, в каком они состоянии и какова их историческая и культурная ценность. Такой учет не может вестись локально. Национальная инфраструктура хранения цифровых копий в стране уже есть – это НЭБ, которая действует

как федеральная государственная информационная система. Однако учет в ней ведется на уровне библиографической записи, а не на уровне конкретного физического экземпляра. Поэтому речь идет не о создании новой системы, а о развитии подсистемы экземплярного учета и координации оцифровки, как в составе НЭБ, так и в составе Реестра научно-технических библиотек и центров НТИ, с интеграцией наук и информации в государственные домены и ресурсы.

Проверить, насколько эффективен предлагаемый подход на практике, можно только через пилотный эксперимент с измерением и отдельной методологической проработкой каждого этапа из всей технологической цепочки. Если такой эксперимент подтвердит существенное повышение эффективности при сохранении нужного качества, потоковая оцифровка вполне может стать одним из инструментов перехода от локальной постраничной оцифровки к промышленному цифровому преобразованию многомиллионных библиотечных фондов.

## Список источников


1. **Technical** Guidelines for Digitizing Cultural Heritage Materials: Third Edition / Federal Agencies Digital Guidelines Initiative (FADGI). Washington, DC, 2023. 129 p. URL: https://digitizationguidelines.gov (дата обращения: 14.08.2026).

2. **Erway R.** Rapid Capture: Faster Throughput in Digitization of Special Collections. Dublin, Ohio: OCLC Research, 2011. DOI 10.25333/C3G34C. URL: https://oclc.org (дата обращения: 08.08.2026).

3. **Zhao J., Huang C.** Technical Issues on the China-US Million Book Digital Library Project // Digital Libraries: International Collaboration and Cross-Fertilization. ICADL 2004. LNCS. Vol. 3334. Springer, 2004. P. 220–226. URL: https://springer.com (дата обращения: 08.08.2026).

4. **Huang C.** Zhejiang University Libraries and CADAL Project: Retrospection & Anticipation // 2014 International Library Cooperation Symposium, 2014. URL: https://jaspul.org (дата обращения: 08.08.2026).

5. **Ambati V., Balakrishnan N., Reddy R., Pratha L., Jawahar C. V.** The Digital Library of India Project: Process, Policies and Architecture // Proceedings of the 2nd International Conference on Digital Libraries (ICDL). New Delhi, 2006. URL: https://iiit.ac.in (дата обращения: 08.08.2026).

6. **Boal G.** Conservation for Digitization at the Wellcome Library // The Book and Paper Group Annual. 2014. Vol. 33. P. 9–16. URL: https://culturalheritage.org (дата обращения: 08.08.2026).

7. **Bartz et al.** v. Anthropic PBC, No. 3:24-cv-05417-WHA, Document 231. Order on Fair Use. United States District Court, Northern District of California, June 23, 2025. URL: https://fingfx.thomsonreuters.com/gfx/legaldocs/jnvwbgqlzpw/ANTHROPIC%20fair%20use.pdf (дата обращения: 08.08.2026).

8. **Сравнение** планетарных сканеров ЭларСкан и Metis EDS Alpha // Elarscan.ru: [официальный сайт]. URL: https://elarscan.ru/comparison/sravnenie-elarskan-a2-vs-metis-eds-alpha/ (дата обращения: 27.08.2026).

9. **Нормы** времени на работы и услуги, выполняемые государственными архивами / Росархив, ВНИИДАД, сост.: З. В. Бушмелева (ответственный исполнитель), В. Е. Соболев, Л. П. Шотина. Москва, 2008. 144 с.

10. **Сколько** зарабатывает сканировщик в городе Москва в 2026 году // Dream Job: сайт отзывов сотрудников о работодателях. URL: https://dreamjob.ru/salary/skanirovschik/city-moskva (дата обращения: 19.08.2026).

11. **Документ-сканер** Epson DS-32000: руководство пользователя и технические характеристики // Epson Russia : [официальный сайт]. URL: https://epson.ru (дата обращения: 14.08.2026).

12. **Документ-сканер** PFU/Ricoh fi-8950: спецификация и эксплуатационные характеристики // Ricoh : [официальный сайт]. URL: https://ricoh.com (дата обращения: 14.08.2026).

13. **Citrome L.** Preserving books by destroying them: a foray into digitization of a professional library // International Journal of Clinical Practice. 2013. Vol. 67, № 10. P. 927–929. DOI 10.1111/ijcp.12316.

14. **Guidelines** for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives / IFLA, UNESCO. 2002. URL: https://repository.ifla.org/items/5160f4c0-739d-44ad-a538-5ab466a962e6 (дата обращения: 25.08.2026).

15. **Kar D. C.** Digital Library of India: An Initiative for the Preservation and Dissemination of the National Heritage and Rare Books and Manuscripts Collection // World Digital Libraries: An International Journal. 2016. Vol. 9, No. 1. DOI 10.18329/09757597/2016/9104. URL: https://bookstore.teri.res.in/e_issue_text_1.php?oj_id=328&sector=792 (дата обращения: 08.08.2026).

16. **Sankar K. P., Ambati V., Pratha L., Jawahar C. V.** Digitizing a Million Books: Challenges for Document Analysis // Document Analysis Systems VII. LNCS. Vol. 3872. Springer, 2006. P. 425–436. URL:https://www.cs.cmu.edu/~vamshi/publications/pramod06Digitizing.pdf (дата обращения: 08.08.2026).

17. **Campagnolo A.** Book Conservation and Digitization: The Challenges of Dialogue and Collaboration. Arc Humanities Press, 2020. 316 p. ISBN 978-1-64189-053-3. URL: https://www.arc-humanities.org/9781641890533/book-conservation-and-digitization/ (дата обращения: 08.08.2026).

18. **Dempsey L., Lavoie B., Malpas C., Connaway L. S., Schonfeld R. C., Shipengrover J. D., Waibel G**. Understanding the Collective Collection: Towards a System-wide Perspective on Library Print Collections. Dublin, Ohio: OCLC Research, 2013. DOI 10.25333/C3GP8R.

19. **Электронная** библиотека РНБ // Российская национальная библиотека : [официальный сайт]. URL: https://nlr.ru/elibrary (дата обращения: 21.08.2026).

20. **Национальная** электронная библиотека (НЭБ) // Фундаментальная библиотека ННГУ : [сайт]. URL: https://lib.unn.ru/elektronnye-resursy/rossijskie-elektronnye-resursy/neb/ (дата обращения: 21.08.2026). Приводится как вторичный источник статистических данных о составе и объеме НЭБ.

21. **Российская** государственная библиотека // Культура. РФ : [официальный портал]. URL: https://www.culture.ru/institutes/15458/rossiiskaya-gosudarstvennaya-biblioteka (дата обращения: 21.08.2026).

22. **О библиотечном** деле: Федеральный закон от 29 декабря 1994 г. № 78-ФЗ (ст. $18^1$ «Национальная электронная библиотека») // Официальный интернет-портал правовой информации. URL: http://publication.pravo.gov.ru (дата обращения: 26.08.2026).

23. **Национальный** информационно-библиотечный центр ЛИБНЕТ : [официальный сайт]. URL: http://www.nilc.ru/ (дата обращения: 26.08.2026).

24. **Шилов В. В.** Однородность библиотечных фондов и функциональные зависимости статистических показателей библиотечной деятельности // Научные и технические библиотеки. 2014. № 5. URL: https://www.gpntb.ru/ntb/ntb/2014/5/ntb_5_5_2014.pdf (дата обращения: 19.08.2026).

25. **О Государственной** системе научно-технической информации: постановление Правительства Российской Федерации от 24 июля 1997 г. № 950 // Официальный интернет-портал правовой информации. URL: http://publication.pravo.gov.ru (дата обращения: 26.08.2026).

26. **Российская** Федерация. Законы. О поддержке развития технологий искусственного интеллекта в Российской Федерации: Федеральный закон от 26 июля 2026 г. № 243-ФЗ // Официальный интернет-портал правовой информации. URL: http://publication.pravo.gov.ru (дата обращения: 14.08.2026).

27. **200 миллионов** книг стали еще доступнее: РГБ запускает национальную ИИ-платформу // Российская государственная библиотека : [официальный сайт]. 03.06.2026. URL: https://www.rsl.ru/news/200-millionov-knig-dostupnee (дата обращения: 26.08.2026).

28. **Российский** сводный каталог по научно-технической литературе // ГПНТБ России : [официальный сайт]. URL: https://www.gpntb.ru/resursy/154-14/5430-rsk-ntl-opisanie.html (дата обращения: 14.08.2026).

29. **Интегрированный** сводный каталог НТИ // ГПНТБ России : [официальный сайт]. URL: https://www.gpntb.ru/proekty/163-deyatelnost/proekty-gpntb/federalnye-programmy/4866-svodnyj-katalog-nti.html (дата обращения: 14.08.2026).

30. **Перечень** поручений по реализации Послания Президента Федеральному Собранию: утв. Президентом Рос. Федерации 30 марта 2024 г. № Пр-616 // Президент России : [официальный сайт]. URL: http://www.kremlin.ru/acts/assignments/orders/73759 (дата обращения: 14.08.2026).

31. **Об утверждении** Порядка учета документов, входящих в состав библиотечного фонда: приказ Министерства культуры Российской Федерации от 08.10.2012 № 1077 (ред. от 02.02.2017). URL: https://www.consultant.ru/document/cons_doc_LAW_146469/ (дата обращения: 08.08.2026).

32. **Гражданский** кодекс Российской Федерации (часть четвертая): ст. 1275 «Свободное использование произведения библиотеками, архивами и образовательными организациями». URL: https://www.consultant.ru/document/cons_doc_LAW_64629/2f837f48edd6ec8026f7b152a88d25f900d16947/ (дата обращения: 08.08.2026).

33. **Гражданский** кодекс Российской Федерации (часть четвертая): ст. 1270 «Исключительное право на произведение». URL: https://www.consultant.ru/document/cons_doc_LAW_64629/dffcf0b87b80ff38f430dc822a0074e76ccd41a0/ (дата обращения: 08.08.2026).

34. **Schaffer A., Oremus W., Tiku N.** Inside an AI start-up's plan to scan and dispose of millions of books // The Washington Post. January 27, 2026. URL: https://www.washingtonpost.com/technology/2026/01/27/anthropic-ai-scan-destroy-books/ (дата обращения: 08.08.2026).

35. **Вахрушев М. В., Гончаров М. В., Колосов К. А.** Что такое датасет и зачем он нужен в библиотеке // Научные и технические библиотеки. 2026;1(7):66-80. https://doi.org/10.33186/1027-3689-2026-7-66-80 С. 66–80. DOI 10.33186/1027-3689-2026-7-66-80.

## References

1. **Technical** Guidelines for Digitizing Cultural Heritage Materials: Third Edition / Federal Agencies Digital Guidelines Initiative (FADGI). Washington, DC, 2023. 129 p. URL: https://digitizationguidelines.gov (data obrashcheniia: 14.08.2026).

2. **Erway R.** Rapid Capture: Faster Throughput in Digitization of Special Collections. Dublin, Ohio: OCLC Research, 2011. DOI 10.25333/C3G34C. URL: https://oclc.org (data obrashcheniia: 08.08.2026).

3. **Zhao J., Huang C.** Technical Issues on the China-US Million Book Digital Library Project // Digital Libraries: International Collaboration and Cross-Fertilization. ICADL 2004. LNCS. Vol. 3334. Springer, 2004. P. 220–226. URL: https://springer.com (data obrashcheniia: 08.08.2026).

4. **Huang C.** Zhejiang University Libraries and CADAL Project: Retrospection & Anticipation // 2014 International Library Cooperation Symposium, 2014. URL: https://jaspul.org (data obrashcheniia: 08.08.2026).

5. **Ambati V., Balakrishnan N., Reddy R., Pratha L., Jawahar C. V.** The Digital Library of India Project: Process, Policies and Architecture // Proceedings of the 2nd International Conference on Digital Libraries (ICDL). New Delhi, 2006. URL: https://iiit.ac.in (data obrashcheniia: 08.08.2026).

6. **Boal G.** Conservation for Digitization at the Wellcome Library // The Book and Paper Group Annual. 2014. Vol. 33. P. 9–16. URL: https://culturalheritage.org (data obrashcheniia: 08.08.2026).

7. **Bartz et al.** v. Anthropic PBC, No. 3:24-cv-05417-WHA, Document 231. Order on Fair Use. United States District Court, Northern District of California, June 23, 2025. URL: https://fingfx.thomsonreuters.com/gfx/legaldocs/jnvwbgqlzpw/ANTHROPIC%20fair%20use.pdf (data obrashcheniia: 08.08.2026).

8. **Sravnenie** planetarny`kh skanerov E`larSkan i Metis EDS Alpha // Elarscan.ru: [ofitcial`ny`i` sai`t]. URL: https://elarscan.ru/comparison/sravnenie-elarskan-a2-vs-metis-eds-alpha/ (data obrashcheniia: 27.08.2026).

9. **Normy`** vremeni na raboty` i uslugi, vy`polniaemy`e gosudarstvenny`mi arhivami / Rosarhiv, VNIIDAD, sost.: Z. V. Bushmeleva (otvetstvenny`i` ispolnitel`), V. E. Sobolev, L. P. Shotina. Moskva, 2008. 144 s.

10. **Skol`ko** zarabaty`vaet skanirovshchik v gorode Moskva v 2026 godu // Dream Job: sai`t otzy`vov sotrudnikov o rabotodateliakh. URL: https://dreamjob.ru/salary/skanirovschik/city-moskva (data obrashcheniia: 19.08.2026).

11. **Dokument-skaner** Epson DS-32000: rukovodstvo pol`zovatelia i tekhnicheskie harakteristiki // Epson Russia : [ofitcial`ny`i` sai`t]. URL: https://epson.ru (data obrashcheniia: 14.08.2026).

12. **Dokument-skaner** PFU/Ricoh fi-8950: spetcifikatciia i e`kspluatatcionny`e harakteristiki // Ricoh : [ofitcial`ny`i` sai`t]. URL: https://ricoh.com (data obrashcheniia: 14.08.2026).

13. **Citrome L.** Preserving books by destroying them: a foray into digitization of a professional library // International Journal of Clinical Practice. 2013. Vol. 67, № 10. P. 927–929. DOI 10.1111/ijcp.12316.

14. **Guidelines** for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives / IFLA, UNESCO. 2002. URL: https://repository.ifla.org/items/5160f4c0-739d-44ad-a538-5ab466a962e6 (data obrashcheniia: 25.08.2026).

15. **Kar D. C.** Digital Library of India: An Initiative for the Preservation and Dissemination of the National Heritage and Rare Books and Manuscripts Collection // World Digital Libraries: An International Journal. 2016. Vol. 9, No. 1. DOI 10.18329/09757597/2016/9104. URL: https://bookstore.teri.res.in/e_issue_text_1.php?oj_id=328&sector=792 (data obrashcheniia: 08.08.2026).

16. **Sankar K. P., Ambati V., Pratha L., Jawahar C. V.** Digitizing a Million Books: Challenges for Document Analysis // Document Analysis Systems VII. LNCS. Vol. 3872. Springer, 2006. P. 425–436. URL:https://www.cs.cmu.edu/~vamshi/publications/pramod06Digitizing.pdf (data obrashcheniia: 08.08.2026).

17. **Campagnolo A.** Book Conservation and Digitization: The Challenges of Dialogue and Collaboration. Arc Humanities Press, 2020. 316 p. ISBN 978-1-64189-053-3. URL: https://www.arc-humanities.org/9781641890533/book-conservation-and-digitization/ (data obrashcheniia: 08.08.2026).

18. **Dempsey L., Lavoie B., Malpas C., Connaway L. S., Schonfeld R. C., Shipengrover J. D., Waibel G**. Understanding the Collective Collection: Towards a System-wide Perspective on Library Print Collections. Dublin, Ohio: OCLC Research, 2013. DOI 10.25333/C3GP8R.

19. **E`lektronnaia** biblioteka RNB // Rossii`skaia natcional`naia biblioteka : [ofitcial`ny`i` sai`t]. URL: https://nlr.ru/elibrary (data obrashcheniia: 21.08.2026).

20. **Natcional`naia** e`lektronnaia biblioteka (NE`B) // Fundamental`naia biblioteka NNGU : [sai`t]. URL: https://lib.unn.ru/elektronnye-resursy/rossijskie-elektronnye-resursy/neb/ (data obrashcheniia: 21.08.2026). Privoditsia kak vtorichny`i` istochnik statisticheskikh danny`kh o sostave i ob``eme NE`B.

21. **Rossii`skaia** gosudarstvennaia biblioteka // Kul`tura. RF : [ofitcial`ny`i` portal]. URL: https://www.culture.ru/institutes/15458/rossiiskaya-gosudarstvennaya-biblioteka (data obrashcheniia: 21.08.2026).

22. **O bibliotechnom** dele: Federal`ny`i` zakon ot 29 dekabria 1994 g. № 78-FZ (st. 18[1] «Natcional`naia e`lektronnaia biblioteka») // Ofitcial`ny`i` internet-portal pravovoi` informatcii. URL: http://publication.pravo.gov.ru (data obrashcheniia: 26.08.2026).

23. **Natcional`ny`i`** informatcionno-bibliotechny`i` centr LIBNET : [ofitcial`ny`i` sai`t]. URL: http://www.nilc.ru/ (data obrashcheniia: 26.08.2026).

24. **Shilov V. V.** Odnorodnost` bibliotechny`kh fondov i funktcional`ny`e zavisimosti statisticheskikh pokazatelei` bibliotechnoi` deiatel`nosti // Nauchny`e i tekhnicheskie biblioteki. 2014. № 5. URL: https://www.gpntb.ru/ntb/ntb/2014/5/ntb_5_5_2014.pdf (data obrashcheniia: 19.08.2026).

25. **O Gosudarstvennoi`** sisteme nauchno-tekhnicheskoi` informatcii: postanovlenie Pravitel`stva Rossii`skoi` Federatcii ot 24 iiulia 1997 g. № 950 // Ofitcial`ny`i` internet-portal pravovoi` informatcii. URL: http://publication.pravo.gov.ru (data obrashcheniia: 26.08.2026).

26. **Rossii`skaia** Federatciia. Zakony`. O podderzhke razvitiia tekhnologii` iskusstvennogo intellekta v Rossii`skoi` Federatcii: Federal`ny`i` zakon ot 26 iiulia 2026 g. № 243-FZ // Ofitcial`ny`i` internet-portal pravovoi` informatcii. URL: http://publication.pravo.gov.ru (data obrashcheniia: 14.08.2026).

27. **200 millionov** knig stali eshche dostupnee: RGB zapuskaet natcional`nuiu II-platformu // Rossii`skaia gosudarstvennaia biblioteka : [ofitcial`ny`i` sai`t]. 03.06.2026. URL: https://www.rsl.ru/news/200-millionov-knig-dostupnee (data obrashcheniia: 26.08.2026).

28. **Rossii`skii`** svodny`i` katalog po nauchno-tekhnicheskoi` literature // GPNTB Rossii : [ofitcial`ny`i` sai`t]. URL: https://www.gpntb.ru/resursy/154-14/5430-rsk-ntl-opisanie.html (data obrashcheniia: 14.08.2026).

29. **Integrirovanny`i`** svodny`i` katalog NTI // GPNTB Rossii : [ofitcial`ny`i` sai`t]. URL: https://www.gpntb.ru/proekty/163-deyatelnost/proekty-gpntb/federalnye-programmy/4866-svodnyj-katalog-nti.html (data obrashcheniia: 14.08.2026).

30. **Perechen`** poruchenii` po realizatcii Poslaniia Prezidenta Federal`nomu Sobraniiu: utv. Prezidentom Ros. Federatcii 30 marta 2024 g. № Pr-616 // Prezident Rossii : [ofitcial`ny`i` sai`t]. URL: http://www.kremlin.ru/acts/assignments/orders/73759 (data obrashcheniia: 14.08.2026).

31. **Ob utverzhdenii** Poriadka ucheta dokumentov, vhodiashchikh v sostav bibliotechnogo fonda: prikaz Ministerstva kul`tury` Rossii`skoi` Federatcii ot 08.10.2012 № 1077 (red. ot 02.02.2017). URL: https://www.consultant.ru/document/cons_doc_LAW_146469/ (data obrashcheniia: 08.08.2026).

32. **Grazhdanskii`** kodeks Rossii`skoi` Federatcii (chast` chetvertaia): st. 1275 «Svobodnoe ispol`zovanie proizvedeniia bibliotekami, arhivami i obrazovatel`ny`mi organizatciiami». URL: https://www.consultant.ru/document/cons_doc_LAW_64629/2f837f48edd6ec8026f7b152a88d25f900d16947/ (data obrashcheniia: 08.08.2026).

33. **Grazhdanskii`** kodeks Rossii`skoi` Federatcii (chast` chetvertaia): st. 1270 «Iscliuchitel`noe pravo na proizvedenie». URL: https://www.consultant.ru/document/cons_doc_LAW_64629/dffcf0b87b80ff38f430dc822a0074e76ccd41a0/ (data obrashcheniia: 08.08.2026).

34. **Schaffer A., Oremus W., Tiku N.** Inside an AI start-up's plan to scan and dispose of millions of books // The Washington Post. January 27, 2026. URL: https://www.washingtonpost.com/technology/2026/01/27/anthropic-ai-scan-destroy-books/ (data obrashcheniia: 08.08.2026).

35. **Vakhrushev M. V., Goncharov M. V., Kolosov K. A.** Chto takoe dataset i zachem on nuzhen v biblioteke // Nauchny`e i tekhnicheskie biblioteki. 2026;1(7):66-80. https://doi.org/10.33186/1027-3689-2026-7-66-80 S. 66–80. DOI 10.33186/1027-3689-2026-7-66-80.

**Информация об авторах / Authors**

**Морозов Иван Владимирович** – научный сотрудник, руководитель службы обеспечения организации управленческой деятельности ГПНТБ России, Москва, Российская Федерация

morozoviv@gpntb.ru

**Михальченкова Наталья Алексеевна** – канд. экон. наук, доктор полит. наук, главный научный сотрудник, генеральный директор ГПНТБ России, Москва, Российская Федерация

MihalchenkovaNA@gpntb.ru

**Ivan V. Morozov** – Researcher, Head, Management Organization Department, Russian National Public Library for Science and Technology, Moscow, Russian Federation

morozoviv@gpntb.ru

**Natalia A. Mikhalchenkova** – Cand. Sc. (Economics), Dr. Sc. (Political Sciences), Chief Researcher, Director General, Russian National Public Library for Science and Technology, Moscow, Russian Federation

MihalchenkovaNA@gpntb.ru