

Современные методы хранения данных в библиотеках: ключевые аспекты и решения

Е. В. Бескаравайная¹, И. А. Митрошин²

^{1, 2}*Библиотека по естественным наукам РАН,
Москва, Российская Федерация*

¹*elenabesk@gmail.com*

²*imitros@gmail.com*

Аннотация. В публикации рассмотрен круг вопросов, возникших перед библиотеками в связи с переводом традиционных фондов в цифровой формат, перечислены технологии для их успешного решения. Хранение научной информации в библиотеках является сложной задачей, требующей учёта множества факторов, таких как устаревание носителей и программного обеспечения, изменение конструкций устройств, совместимость форматов, безопасность данных и доступ к информации.

Подняты вопросы выбора документов для оцифровки в соответствии со спецификой библиотек и потребностями пользователей, предложены варианты подготовки данных, отобранных на хранение. Одна из основных проблем, затронутых в статье, связана с трудностями миграции данных, их потерей или повреждением в процессе переноса, неправильной конфигурации или недостаточной проверки данных перед миграцией. Сделаны выводы о перспективности поиска технологий с элементами машинного анализа и алгоритмами машинного обучения, направленными на различные сферы библиотечной деятельности, в том числе на хранение данных. Указана необходимость интеграции различных служб и ведомств в процессе перевода данных в цифровой формат.

Ключевые слова: научные и технические библиотеки, цифровизация, хранение данных, миграция данных, машинный анализ

Для цитирования: Бескаравайная Е. В., Митрошин И. А. Современные методы хранения данных в библиотеках: ключевые аспекты и решения // Научные и технические библиотеки. 2024. № 8. С. 78–97. <https://doi.org/10.33186/1027-3689-2024-8-78-97>

DIGITAL INFORMATION RESOURCES

UDC 004:021

<https://doi.org/10.33186/1027-3689-2024-8-78-97>

The modern methods of data preservation in libraries. The key aspects and solutions

Elena V. Beskaravainaya¹ and Ivan A. Mitroshin²

^{1,2}*Library for Natural Sciences, Russian Academy of Sciences,
Moscow, Russian Federation*

¹*elenabesk@gmail.com*

²*imitros@gmail.com*

Abstract. In this publication, we examined the range of issues that the libraries face during the transfer of traditional collections into digital format, and provided examples of efficient technological solutions.

Preserving scientific information in libraries is a complex task encompassing many factors, such as: media obsolescence, changes in device designs, software obsolescence and updates, format compatibility, data security and access to information.

The issues of selecting documents for digitization in accordance with the specific requirements of the library and user needs are discussed, and several options for preparing selected data for preservation are proposed.

One of the main problems is related to the difficulties of data migration, data loss or damage during the migration process, incorrect configuration or insufficient data verification before migration. The conclusions are made on the prospects for technologies with machine analysis and machine learning algorithms designed for various library services, including data preservation. The need for integration of various library departments in the process of transferring data into digital formats is emphasized.

Keywords: scientific and technical libraries, digitalization, data preservation, data migration, machine analysis

Cite: Beskaravainaya E. V., Mitroshin I. A. The modern methods of data preservation in libraries. The key aspects and solutions // Scientific and technical libraries. 2024. No. 8, pp. 78–97. <https://doi.org/10.33186/1027-3689-2024-8-78-97>

В современном мире научная информация является одним из наиболее ценных для человечества ресурсов: каждое научное открытие или полученные в ходе исследований данные потенциально способны изменить наше восприятие мира и содействовать решению мировых проблем. Так как объём данных постоянно растёт, а достижения в науке во многом основаны на предыдущих исследованиях, представленных в научных документах, обеспечение долгосрочного хранения информации – критически важная задача, которую предстоит решать современной библиотеке.

Актуальность темы обусловлена непрерывным развитием технологий (в том числе и информационных) и программного обеспечения, требующих постоянной адаптации к условиям функционирования библиотек с целью оптимизации их услуг.

Цель нашей работы – рассмотрение основных методов и способов хранения, отбора и обработки информации, обсуждение роли информационных технологий в обеспечении хранения научной информации.

В качестве метода исследования мы применяли метод наблюдения, основанный на собственном опыте многолетней работы в академической библиотеке, и метод обобщения, основанный на изучении публикаций сотрудников отечественных и зарубежных библиотек и специалистов, обслуживающих программное обеспечение в них. Публикации собирались с помощью поисковой системы Google Scholar и информационно-аналитического портала eLIBRARY.RU, по ключевым словам: **«хранение в библиотеках»**, «цифровое архивирование в библиотеках», «цифровая сохранность в библиотеках», «хранение данных в репозиториях», «цифровизация фондов библиотек», «миграция данных», «применение интеллектуального анализа в библиотеках» на русском и английском языках. Критериями отбора литературы являлись: 1) хронологический период не старше 20 лет; 2) применение результатов описанных исследований на практике в библиотеках или информационных центрах.

Хотя в существующей литературе предлагаются различные составляющие организации новых сервисов, учитывающие различия в типах библиотек, целевой аудитории и финансовых возможностях, можно выделить некоторые общие предварительные задачи и проблемы, встающие перед библиотекарями на начальном этапе организации хранения. Это определение понятия «данные» применительно к биб-

лиотечному делу, стратегия сбора литературы для хранения, оценки наборов данных при отборе на долговременное хранение.

Сохранение данных, включая цифровой контент, всё ещё является относительно молодой областью, непрерывно развивающейся и требующей *выработки единых терминов и определений*. Анализ понятия «данные» применительно к библиотечному делу провели M. Zhan & G. Widén (2019) [1] по статьям из баз Library, Information Science & Technology Abstract (EBSCO) и Web of Science, Scimago Journal & Country Ranking. Они пришли к выводу о существовании двух способов определения больших данных в библиотечном деле: ориентированных на данные (большой, быстро увеличивающийся объём) и ориентированных на возможности (технология обработки). Поскольку G. Bellinger с соавторами [2] рассматривают информацию как «данные, которые обрабатываются для использования», можно предположить, что данными может являться и информация, расширяя таким образом область применения этого понятия. Объединив эти два определения, приходим к выводу, *что в библиотечном деле это данные или информация, подлежащие обработке с помощью развитых технологий*.

Одним из вопросов является *стратегия сбора литературы*. Для статей из академических баз данных и журналов J. Frizzo-Barker [3] предлагает метод перекрёстного поиска литературы, когда статьи проверяются на наличие в разных источниках для обеспечения полноты поиска. Briner R. с соавторами советует использовать систематические обзоры и отталкиваться от конкретной тематики [4]. Основываясь на принципах «находимость», «доступность», «взаимодействие» и «повторное использование», M. Martone с коллегами [5] уделяют особое внимание расширению возможностей находить и использовать данные *автоматически*, а также поддерживать их повторное использование через применение различных утилит. При этом критерии отбора данных [6] для обеспечения научных исследований остаются неизменными: типология, статистика использования, библиометрическая оценка, экспертная система комплектования и др. [7–9].

Методы *оценки наборов данных при отборе на долговременное хранение* – ещё одна насущная задача. Австралийская национальная служба предлагает семь общих критериев: соответствие миссии, научная или историческая ценность, уникальность, потенциал для перераспределения, невозпроизводимость, экономическое обоснование, пол-

нота документации [10]. Ключевыми концепциями Службы данных Великобритании являются: соответствие компетенции стандартам государства, научная или историческая ценность, новые источники или типы данных, международная ценность, уникальность/риск потери, удобство использования/перераспределение, эксплуатационные преимущества, репликация [11]. Digital Curation Centre (DCC) в своём руководстве по отбору данных предлагает включить: потенциальные цели для повторного использования, риски, связанные с соблюдением законодательства или политики, долгосрочная ценность, завершённость оценки данных [12]. Учёные из США, Испании, Нидерландов на начальном этапе выделяют четыре составляющих для отбора метаданных: «находимость», «доступность», «взаимодействие» и «повторное использование» [13]. Итак, как отмечает большинство исследователей, основными критериями являются *целостность*, *воспроизводимость* (наличие данных, подтверждающих результаты исследования), *возможность повторного использования* (доступность данных для обмена с другими пользователями) [14].

На данный момент общей концепции по сохранению цифровых данных, которая включала бы сразу все проблемные точки: финансовые и кадровые вопросы, вопросы интеллектуальной собственности, соответствие стандартам, стратегию хранения носителя, сертифицирование, аудит, оценку рисков и пр., не существует. *Политику сохранения цифровых данных* библиотеки чаще всего разрабатывают с помощью служб курирования, например, Digital Curation Centre, Jisc и др. (Великобритания), Roper Center for Public Opinion Research и др. (США), LinguaManga и др. (Китай), Библиотека для открытой науки (Россия); либо используют собственную службу, направленную на эту деятельность, как в Библиотечной системе Массачусетского технологического института, Национальном научном фонде в Великобритании, Ассоциации европейских исследовательских библиотек LIBER в Нидерландах.

Привлекая инновационные технологии для решения собственных задач и меняющихся потребностей пользователей, библиотекам необходимо решить вопросы хранения как традиционных фондов, так и цифровых материалов.

В большинстве библиотек основной объём информации пока представлен на бумаге: книги, журналы, газеты, рукописи и другие печатные издания. На первый взгляд, надёжность хранения информации

на бумажных и фотографических носителях не вызывает сомнений. Однако следует учитывать ряд важных аспектов. Во-первых, качество бумаги может варьироваться в зависимости от времени её изготовления: бумага XIX в. и XXI в. – это два разных типа бумаги [15]. Во-вторых, даже в пределах одного столетия бумага может быть разной: «финской», «писчей» и т. д. Кроме того, материалы представления информации на бумаге также меняются: чернила, тушь, карандаши, пишущие машинки, лазерные принтеры и т. д. Утверждение о том, что бумага способна храниться тысячу лет, является достаточно условным. Недостатками традиционных фондов следует считать физический износ, зависимость от пожаров и влажности. Таким образом, *методом хранения традиционных документов является консервация*, направленная на обеспечение долговременной защиты от неблагоприятных воздействий окружающей среды (ГОСТ 7.50-2002 «СИБИД. Консервация документов. Общие требования»). Сложность консервации заключается в нарастающем объёме материала и ежедневном контроле за внешними условиями и частично решается оцифровкой.

Итак, вторым методом хранения библиотечных фондов является *перевод документов в цифровой формат*, позволяющий представить ресурсы в более компактном виде, сделать их удобнее для автоматического поиска, снять вопрос физической доступности.

При этом актуальной задачей становится выбор документов для первоочередного перевода в цифровой формат. Оцифровка библиотечных фондов ГПНТБ России ведётся на базе специально созданного для этих целей подразделения «Центр сканирования» и включает все категории документов, подлежащих оцифровке. Национальная библиотека Беларуси заключила договор с Национальным центром интеллектуальной собственности на передачу авторских неимущественных прав для оцифровки музыкальных коллекций, позволяющий разрешить одну из сложнейших проблем – поиск автора произведения. Приоритетными документами для оцифровки в Вологодской областной универсальной научной библиотеке, Архангельской областной научной библиотеке, национальной библиотеке Республики Саха являются издания, представляющие интерес для фольклористов, языковедов, краеведов. Центральная научная сельскохозяйственная библиотека при создании цифровых документов отдаёт предпочтение книжным памятникам и наиболее востребованным экземплярам изданий в плохом физическом

состоянии. Российская государственная библиотека искусств переводит в цифровой формат гравюры, театральные эскизы, открытки, уникальные коллекции оригинальных фотографий, театральных программ и др. В Библиотеке по естественным наукам РАН, кроме планового перевода фондов в цифровой формат, пользователи могут предложить для оцифровки интересное их издание.

Среди крупных отечественных проектов по цифровизации, предполагающих тесное взаимодействие различных организаций, следует упомянуть Национальную электронную библиотеку – НЭБ (<https://rusneb.ru/>), электронную библиотеку «Научное наследие России» (<http://e-heritage.ru/Catalog/About>); Научную электронную библиотеку eLIBRARY.RU (<https://elibrary.ru/defaultx.asp>) и др., источниками комплектования которых являются библиотечные, архивные, музейные фонды академических организаций.

При переводе фондов в цифровой формат появляются совершенно новые задачи по обеспечению хранения уже цифровых фондов [16]:

- сохранение подлинности и цельности данных в течение всего периода хранения;

- поддержание возможности интерпретации (воспроизведения формата цифровой информации) в течение всего периода хранения;

- обеспечение надёжности носителя или системы хранения цифровой информации.

Хранение цифровой информации в библиотеках подразумевает использование физических носителей: перфокарт, магнитных лент, магнитных дисков (жёсткие диски HDD и дискеты), оптических накопителей (CD, DVD, Blu-ray Discs) твёрдотельных накопителей (SSD, флеш-память) и технических устройств, обслуживающих их работу. Для применения в библиотеках мы рекомендуем использовать диски SSD и HDD. Со временем их скорость и производительность уменьшаются, а гарантия на работу составляет не более 10 лет. Но, на наш взгляд, этот выбор пока оптимален по стоимости и надёжности.

На данный момент в мире нет опыта действительно долгого сохранения цифровых данных, отсюда некоторая иллюзорность предсказания спектра возможных проблем. Именно поэтому одним из методов хранения и цифровых, и бумажных фондов является *консервация* – сохранение материалов в оригинальных форматах и на оригинальных носителях (включая микрофильмы, микрофиши, оптические диски).

Нерешённой проблемой, с которой сталкиваются все библиотеки при цифровизации, является обилие разнообразных форматов данных (текст, изображения, видео, статистические таблицы, ценные ресурсы), подразумевающие различные подходы к их обработке и архивированию [17–19]. Данные, записанные несколько десятков лет назад на цифровые носители, даже при надлежащем хранении становятся недоступными не только из-за устаревания носителей, но и из-за способов записи, а также прекращения поддержки со стороны производителей программного обеспечения и операционных систем. Методом хранения, позволяющим сгладить эту проблему, является *эмуляция* – создание совокупности программных и технических средств, позволяющих воспроизводить функциональность исходной аппаратно-программной среды.

Опираясь на собственный опыт, хочется обратить внимание на возможную потерю информации как при хранении, так и при переносе с одного носителя на другой, в связи с чем важнейшими методами сохранения цифровых данных становятся *резервное копирование* (сохранение информации на отдельном сервере) и *репликация* (хранение данных в нескольких местах).

В качестве варианта *цифрового хранения* для больших тематических коллекций [20–22] наиболее распространённым и приемлемым, в том числе для российских библиотек [23], на сегодняшний день являются репозитории, такие как OpenDOAR (глобальный каталог репозитория открытого доступа – <https://v2.sherpa.ac.uk/opensdoar/>), bioRxiv.org (бесплатный электронный архив научных статей и препринтов по биологии – <https://www.biorxiv.org/>), The World Bank Open Knowledge Repository (официальное хранилище результатов исследований Всемирного банка – <https://openknowledge.worldbank.org/home>), НОРА (Национальный агрегатор открытых репозиториях российских университетов – <https://openrepository.ru/>) и мн. др. Наиболее полная информация и уникальные данные, собранные в одном месте, актуальны для пользователей, работающих в узких областях или имеющих специфические потребности. Посетителям обеспечивается экономия времени, а библиотеки получают возможность осуществлять контроль за данными, отслеживать использование материалов, определять потребности читателей, оптимизировать процессы приобретения новых изданий и т. д. Кроме того, поддержка репозиториях позволяет перераспределить фонд библиотек и перенести фокус внимания с хранения

коллекций и покупки материалов на курирование контента, в том числе открытого доступа [24–26].

В будущем именно библиотечные репозитории с качественно отобранными и структурированными данными могут стать *основой для машинного анализа* – процесса исследования и интерпретации данных с помощью специализированного программного обеспечения и алгоритмов, которые изначально применялись для анализа текста в поисковых системах, чат-ботах и при переводах. Уже сегодня отечественные и зарубежные авторы, описывая аналитику данных в библиотечном контексте при переходе к цифровым ресурсам, цифровому хранению, онлайн-сервисам с акцентом на открытый доступ, говорят о новых возможностях по улучшению функционирования библиотек при использовании искусственного интеллекта для сбора данных и их дальнейшей обработки [27–32]. Мы согласны с тем, что машинный анализ является одним из перспективных направлений в библиотеках, где содержатся и обрабатываются большие и разнообразные объёмы данных, включающие в себя информацию о публикациях, авторах, темах исследований, патентах, экспериментальных данных, пользовательской активности и других аспектах научной деятельности. Опираясь на хранимую в библиотеке информацию, машинный анализ может использоваться для разработки, например, собственных рекомендательных систем, интегрированных в «личные кабинеты учёных» [33; 34], основываясь на анализе доступных ресурсов, поведении пользователей, их предпочтениях и предыдущих запросах [35]; изучать и прогнозировать спрос на определённые книги, журналы и др. материалы; анализировать тренды в использовании фондов библиотеки для нахождения наиболее популярных тем и направлений исследований; проводить анализ качества переводов [36–41].

Ещё одним методом хранения, позволяющим решить проблему износа и устаревания, является *регулярное обновление* – перенос информации с одного физического носителя на идентичный. Несмотря на то, что обновление не считается долговременным способом архивирования, оно позволяет сохранить информацию для более сложного процесса – *миграции данных* (преобразование данных на другие носители или операционную систему) [42; 43]. Миграция данных в библиотеках – самый сложный процесс хранения: требуется учитывать не только различие в технологиях (могут иметься цифровые копии, полученные на

разных этапах развития техники, запакованные в различные форматы), но и специфику перехода от одной АБИС к другой (перенос каталогов и коллекций книг, журналов, электронных ресурсов, патентов, информации о читателях, электронных ресурсов и других данных, необходимых для работы библиотеки). Миграция данных, начиная от этапа планирования и заканчивая проверкой целостности сохранённой информации, потребует установки программного обеспечения и привлечения в работу библиотек программистов-специалистов.

Перевод традиционных фондов в цифровые и обеспечение их сохранности затрагивает принципиальные вопросы *безопасности*, связанные с управлением правами с течением времени. Например, каков механизм предоставления доступа для различных групп пользователей при взаимодействии с метаданными архива, как будут обновляться права доступа при изменении статуса авторских прав или уровня безопасности материала [44–46]. Эталонные модели архивирования как стандарты цифровых архивов с защитой цифровых сведений, устройств и ресурсов были разработаны ещё в начале XXI в. (CCSDS 2001 [47], OAIS [48] и др.) и сегодня используются в большинстве крупных библиотек Австралии, Великобритании, Нидерландов и США. На их основе существуют крупные проекты по сохранности данных [49]: Cedars – консорциум университетских исследовательских библиотек, Kulturaw3 – Королевская библиотека Швеции, NEDLIB – сеть европейских депозитных библиотек, LOCKSS – библиотеки Стэнфордского университета, PANDORA – Национальная библиотека Австралии, SPAR – Национальная библиотека Франции [50], EVA – Национальная библиотека Финляндии при Хельсинкском университете [51].

Серьёзную задачу для сбора, анализа, визуализации, обработки данных и их воспроизводимости [52] в библиотечных системах представляет собой *разработка программного обеспечения*, включающего в себя как исходный код, так и исполняемые файлы, используемые в процессе сохранности данных. Примерами поставщиков этой категории [53], обеспечивающих программное сохранение, являются CLOCKSS, Internet Archive, HathiTrust и Portico. Кроме того, существуют сторонние платформы сохранения с собственными технологиями (APTrust, Archivematica, Arkivum, Islandora, LIBNOVA, MetaArchive, Samvera и Preservica), которые используются отдельными организациями, и, как правило, не бесплатно.

Выводы

В настоящее время перед библиотеками стоят четыре *глобальные задачи по сохранению фондов*:

1. Создание условий для сохранения традиционного фонда.
2. Перевод документов в цифровой формат.
3. Поддержка сохранности документов в любом формате.
4. Обеспечение безопасности данных на всех этапах хранения.

Основными этапами при *подготовке к хранению данных* любых форматов должны оставаться: фильтрация данных (удаление шума, дубликатов, ошибок, пропусков и т. д.); масштабирование данных для облегчения сравнения (нормализация и стандартизация); преобразование и адаптация к конкретному аналитическому инструменту; агрегирование и интеграция данных, объединение информации из разных источников и преобразование её в единую систему.

Следует отметить, что невозможно выбрать и использовать какой-то один метод для хранения, они не являются взаимозаменяемыми и требуют совокупного применения. *Методами хранения информации* в библиотеках и информационных центрах, объединяющими программные средства и технологии, являются:

консервация,
перевод документов в цифровой формат,
эмуляция,
резервное копирование,
репликация,
регулярное обновление,
миграции данных,
облачное хранение.

Одной из основных проблем, связанных с переносом данных и их хранением, является выбор носителя для хранения цифровых данных. Следует учитывать его стабильность, надёжность и совместимость с существующими системами хранения.

Разработка процессов миграции данных – важная задача, включающая в себя поиск специальных алгоритмов, которые позволяли бы не только эффективно переносить данные на цифровой носитель, но и сохранять их целостность после переноса.

Хочется подчеркнуть, что затронутые в статье вопросы разработки и проверки новых алгоритмов хранения, их стандартизации должны решаться при взаимодействии библиотек с архивами, музеями, научными лабораториями, технологическими и исследовательскими компаниями. Такая работа потребует экономических, юридических, финансовых, организационных усилий не только со стороны библиотек и институтов памяти.

Список источников

1. **Zhan M., & Widén G.** Understanding big data in librarianship // Journal of Librarianship and Information Science. 2019. V. 51. № 2. Pp. 561–576. DOI 10.1177/0961000617742451.
2. **Bellinger G., Castro D., Mills A.** Data, Information, Knowledge, and Wisdom. 2004. URL: <http://courseweb.ischool.illinois.edu/~katewil/spring2011-502/502andotherreadings/bellingeronackoffdatainfknowwisdom.pdf>.
3. **Frizzo-Barker J., Chow-White P. A., Mozafari M.** An empirical study of the rise of big data in business scholarship // International Journal of Information Management. 2006. V. 36. № 3. Pp. 403–413.
4. **Briner R. B., Denyer D., Rousseau D. M.** Evidence-based management: concept cleanup time? // Academy of management perspectives. 2009. V. 23. № 4. P. 19–32.
5. **Martone M. E.** FORCE11: Building the Future for Research Communications and e-Scholarship // BioScience. 2015. V. 65. i. 7. P. 635. DOI 10.1093/biosci/biv095.
6. **Г. М. Вихрева, О. П. Федотова.** Отбор периодических изданий в фонды крупной универсальной научной библиотеки: конспект лекций для слушателей учреждений дополнительного профессионального образования (специальность 071201 «Библиотечно-информационная деятельность») / Сибирское отделение Российской академии наук, Государственная публичная научно-техническая библиотека СО РАН. Новосибирск : Государственная публичная научно-техническая библиотека СО РАН, 2010. 103 с. ISBN 978-5-94560-192-5.
7. **Гуреев В. Н., Мазов Н. А.** Модели и критерии отбора изданий в фонд научной библиотеки // Научные и технические библиотеки. 2015. № 7. С. 31–50.
8. **Bocharova E. N., Kochukova E. V.** The modern concept of selecting publications for the research library collections. Scientific and Technical Libraries. 2019. 4. С. 3–16. (In Russ.) DOI 10.33186/1027-3689-2019-4-3-16.
9. **Харыбина Т. Н., Бескаравайная Е. В., Митрошин И. А.** Модель исследования информационной значимости иностранных журналов // Наукометрия: методология, инструменты,

практическое применение: сб. научных статей / под ред. А. И. Груши. Минск, 2018. С. 257–276.

10. **Whyte A., Wilson A.** How to Appraise and Select Research Data for Curation. DCC How-to Guides. Edinburgh: Digital Curation Centre. 2010. URL: dcc.ac.uk/resources/how-guides/appraise-select-data.

11. **Data** management costing tool and checklist, version 3, UK Data Service. UK Data Service (UKDS). 2015. URL: ukdataservice.ac.uk/manage-data/plan/costing.

12. **DCC.** Five steps to decide what data to keep: checklist for appraising research data // Edinburgh: Digital Curation Centre. 2014. V. 1. URL: dcc.ac.uk/resources/how-guides.

13. **Tenopir C., Rice N. M., Allard S., Baird L., Borycz J., Christian L. et al.** Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide // PLoS ONE. 2020. V. 15. № 3. e0229003. DOI 10.1371/journal.pone.0229003.

14. **Beagrie N.** What to Keep: A Jisc research data study. 2019. URL: <https://apo.org.au/node/227001>.

15. **Залаев Г. З., Цветкова В. А.** Некоторые вопросы хранения цифровой информации // НТИ-2022. Научная информация в современном мире: глобальные вызовы и национальные приоритеты : материалы 10-й научной конференции с международным участием, посвящённой 70-летию ВИНИТИ РАН, Москва, 25–26 октября 2022 г. Москва : Всероссийский институт научной и технической информации РАН, 2022. С. 73–75. DOI 10.36535/2022-9785945770829-14.

16. **Залаев Г. З., Каленов Н. Е., Цветкова В. А.** Оцифровка документов в научных архивах и библиотеках: вопросы и ответы // НТИ. Сер. 1: Организация и методика информационной работы. 2016. № 2. С. 14–21.

17. **Soohyung J., Hofman D., Youngseek K.** Investigation of challenges in academic institutional repositories: A survey of academic librarians // Library Hi Tech. 2019. V. 37. № 3. Pp. 525–548.

18. **Tait E.** Libraries for the future: the role of IT utilities in the transformation of academic libraries // Palgrave Communications. 2016. V. 2. 16070. DOI 10.1057/palcomms.2016.70.

19. **Roche D. G., Lanfear R., Binning S. A., Haff T. M., Schwanz L. E., Cain K. E.** Troubleshooting Public Data Archiving: Suggestions to Increase Participation // PLoS Biology. 2014. V. 12. № 1. e1001779. DOI 10.1371/journal.pbio.1001779.

20. **Salomon J. A.** Stanford University School of Medicine, Stanford, CA, and accepted by the Editorial Board October 18 // PNAS. 2021. V. 118. e2111454118.

21. **Xueqin Guo, Fengzhen Chen, Fei Gao, Ling Li, Ke Liu, Lijin You, Cong Hua, Fan Yang, Wanliang Liu, Chunhua Peng, Lina Wang, Xiaoxia Yang, Feiyu Zhou, Jiawei Tong, Jia Cai, Zhiyong Li, Bo Wan, Lei Zhang, Tao Yang, Minwen Zhang, Linlin Yang, Yawen Yang, Wenjun Zeng, Bo Wang, Xiaofeng Wei, Xun Xu.** CNSA: a data repository for archiving omics data, Database. 2020, 2020:baaa055. DOI 10.1093/database/baaa055.

22. **Fevola C., Görgen C.** The mathematical research-data repository MathRepo // arXiv preprint arXiv:2202.04022. 2022. URL: <https://arxiv.org/abs/2202.04022>.

23. **Рождественская М. Ю.** Репозиторий как реализация идей открытого доступа к научным публикациям: подходы к классификации // Библиосфера. № 2. 2015. Pp. 86–94.

24. **Kouper I.** Professional participation in digital curation // *Library & Information Science Research*. 2016. Т. 38. № 3. С. 212–223.
25. **Kowalczyk S. T.** Digital curation for libraries and archives // Bloomsbury Publishing USA. 2018. ISBN 978-1610696319.
26. **Jørn Nielsen H., Hjørland B.** Curating research data: the potential roles of libraries and information professionals // *Journal of documentation*. 2014. Т. 70. № 2. С. 221–240.
27. **Tiwari Amit.** Data analytics in libraries and information centres. 2018. URL: https://www.researchgate.net/publication/327389981_data_analytics_in_libraries_and_information_centres.
28. **Tercan H., Meisen T.** Machine learning and deep learning based predictive quality in manufacturing: a systematic review // *Journal of Intelligent Manufacturing*. 2022. V. 33. Pp. 1879–1905 (2022). DOI 10.1007/s10845-022-01963-8.
29. **Biswas A.** Application of Data Analytics for Mapping of Library System and Services of Oxford University Library. 2023. URL: https://www.researchgate.net/publication/375447227_Application_of_Data_Analytics_for_Mapping_of_Library_System_and_Services_of_Oxford_University_Library.
30. **Hsin-liang Chen, Philip Doty, Carol Mollman, Xi Niu, Jen-chien Yu, Tao Zhang.** Library assessment and data analytics in the big data era: Practice and policies // *Proceedings of the Association for Information Science and Technology*. 2020. V. 52. I. 1. Pp. 1–4. DOI 10.1002/pra2.2015.14505201002.
31. **Preza Díaz J. L.** Data Science and Analytics in Libraries // *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*. 2017. V. 70. P. 244. DOI 10.31263/voebm.v70i2.1796.
32. **Редькина Н. С.** Векторы развития научных библиотек: обзор ключевых докладов Всемирного конгресса ИФЛА 2019 г. // *Библиосфера*. 2020. № 2. С. 71–81. DOI 10.20913/1815-3186-2020-2-71-81.
33. **Кветкина Е. А.** Личный кабинет пользователя как виртуальный сервис информационного сопровождения библиотекой научно-исследовательской деятельности. Репозиторий ЦНБ НАН Беларуси. 2023. С. 136–143.
34. **Горбич Л. Г.** Информационная система «Web-кабинет учёного» как интерактивная электронная библиотека // *Библиотеки вузов Урала: проблемы и опыт работы*. 2015. В. 14. № 14. С. 70–72.
35. **Митрошин И. А.** Основные принципы развития сайта научной библиотеки // *Научные и технические библиотеки*. 2020. № 11. С. 165–184. DOI 10.33186/1027-3689-2020-11-165-184.
36. **Нуриев В. А., Егорова А. Ю.** Методы оценки качества машинного перевода: современное состояние, Информатика и её применение. 2021. Т. 15. В. 2. С. 104–111. DOI 10.14357/19922264210215.
37. **Конев К. А.** Использование методов машинного обучения в задачах принятия решений при обеспечении качества в приборостроении // *Экономика. Информатика*. 2022. V. 49. № 4. Pp. 820–832.

38. **Веретенников И. С., Карташев Е. А., Царегородцев А. Л.** Оценка качества классификации текстовых материалов с использованием алгоритма машинного обучения «Случайный лес» // Известия Алтайского государственного университета. 2017. № 4 (96). Pp. 78–83.
39. **Kleijnen Jack & van Groenendaal Willem.** Measuring The Quality Of Publications: New Methodology And Case Study // Inf. Process. Manage. 2000. V. 36. Pp. 551–570. DOI 10.2139/ssrn.247676.
40. **Mike Thelwall.** Can the quality of published academic journal articles be assessed with machine learning? Quantitative Science Studies. 2022. V 3. № 1. Pp. 208–226. DOI 10.1162/qss_a_00185.
41. **Zhou J., Gandomi A. H., Chen F., Holzinger A.** Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics // Electronics. 2021. V. 10. P. 593. DOI 10.3390/electronics10050593.
42. **Becker C., Kulovits H., Guttenbrunner M. et al.** Systematic planning for digital preservation: evaluating potential strategies and building preservation plans // International Journal on Digital Libraries. 2009. V.10. Pp. 133–157 DOI:10.1007/s00799-009-0057-1.
43. **Van der Hoeven J., van Wijngaarden H.** Modular emulation as a long-term preservation strategy for digital objects // In: 5th International Web Archiving Workshop (IWA05). Vienna, Austria. 2005. URL: <https://api.semanticscholar.org/CorpusID:9066566>.
44. **Бегишев И. Р.** Правовые средства обеспечения безопасности цифровых архивов в условиях внедрения технологий искусственного интеллекта // Вестник Юридического института МИИТ. 2021. № 2 (34). С. 108–116.
45. **El Kafhali S., El Mir I., Hanini M.** Security Threats, Defense Mechanisms, Challenges, and Future Directions in Cloud Computing // Archives of Computational Methods in Engineering. 2022. V. 29. 223–246. DOI 10.1007/s11831-021-09573-y.
46. **Alcaraz C., Lopez J.** Digital Twin: A Comprehensive Survey of Security Threats // IEEE Communications Surveys & Tutorials. 2022. Vol. 24. DOI 10.1109/COMST.2022.3171465.
47. **Yeh P. S., Moury G. A., Armbruster P.** CCSDS data compression recommendations: development and status // Applications of Digital Image Processing XXV // SPIE. 2002. V. 4790. Pp. 302–313.
48. **Zierau E., Schultz M.** Creating a framework for applying OAIS to distributed digital preservation // Proceedings of the 10th International Conference on Preservation of Digital Objects. 2013. Pp. 78–83.
49. **Hodge G.** Archiving and Preservation in Electronic Libraries // Electronic Information Management for PFP Nations. 2002. URL: https://www.researchgate.net/publication/235146965_Archiving_and_Preservation_in_Electronic_Libraries.
50. **Ledoux T.** Long-term preservation at the National Library of France (BnF): Scalable Preservation and Archiving Repository (SPAR) // International Preservation. 2012. Pp. 2. URL: https://irods.org/uploads/2014/06/BnF_2014_06_19_iRodsUG_BnF-ADM-2014-061850-01-p2.pdf.

51. **Lounamaa K.** EVA – The Acquisition and Archiving of Electronic Network Publications In Finland. 1999. URL: <https://www.ercim.eu/publication/ws-proceedings/DELOS6/eva.pdf>.
52. **Al Noamany Y., Borghi J. A.** Towards computational reproducibility: researcher perspectives on the use and sharing of software // *PeerJ Computer Science*. 2018. № 4. e163 DOI 10.7717/peerj-cs.163.
53. **Housewright R., Schonfeld R. C., Wulfson K.** US faculty survey 2012 // *Ithaka S + R*. 2013. Pp. 45–80. DOI 10.18665/sr.22502.

References

1. **Zhan M., & Widén G.** Understanding big data in librarianship // *Journal of Librarianship and Information Science*. 2019. V. 51. № 2. Pp. 561–576. DOI 10.1177/0961000617742451.
2. **Bellinger G., Castro D., Mills A.** *Data, Information, Knowledge, and Wisdom*. 2004. URL: <http://courseweb.ischool.illinois.edu/~katewill/spring2011-502/502andotherreadings/bellingeronackoffdatainfknowwisdom.pdf>.
3. **Frizzo-Barker J., Chow-White P. A., Mozafari M.** An empirical study of the rise of big data in business scholarship // *International Journal of Information Management*. 2006. V. 36. № 3. Pp. 403–413.
4. **Briner R. B., Denyer D., Rousseau D. M.** Evidence-based management: concept cleanup time? // *Academy of management perspectives*. 2009. V. 23. № 4. P. 19–32.
5. **Martone M. E.** FORCE11: Building the Future for Research Communications and e-Scholarship // *BioScience*. 2015. V. 65. i. 7. P. 635. DOI 10.1093/biosci/biv095.
6. **G. M. Vikhreva, O. P. Fedotova.** Otbor periodicheskikh izdaniï v fondy` krupnoi` universal`noi` nauchnoi` biblioteki: konspekt lektcii` dlia slushatelei` uchrezhdenii` dopolnitel`nogo professional`nogo obrazovaniia (spetsial`nost` 071201 «Bibliotechno-informatcionnaia deiatel`nost`») / Sibirskoe otdelenie Rossii`skoi` akademii nauk, Gosudarstvennaia publichnaia nauchno-tekhnicheskaia biblioteka SO RAN. Novosibirsk : Gosudarstvennaia publichnaia nauchno-tekhnicheskaia biblioteka SO RAN, 2010. 103 s. ISBN 978-5-94560-192-5.
7. **Gureev V. N., Mazov N. A.** Modeli i kriterii otbora izdaniï v fond nauchnoi` biblioteki // *Nauchny`e i tekhnicheskije biblioteki*. 2015. № 7. S. 31–50.
8. **Bocharova E. N., Kochukova E. V.** The modern concept of selecting publications for the research library collections. *Scientific and Technical Libraries*. 2019. 4. C. 3–16. (In Russ.) DOI 10.33186/1027-3689-2019-4-3-16.
9. **Hary`bina T. N., Beskaravai`naia E. V., Mitroshin I. A.** Model` issledovaniia informatcionnoi` znachimosti inostranny`kh zhurnalov // *Naukometriia: metodologiya, instrumenty`, prakticheskoe primeneniie: sb. nauchny`kh statei` / pod red. A. I. Grushi*. Minsk, 2018. S. 257–276.
10. **Whyte A., Wilson A.** *How to Appraise and Select Research Data for Curation*. DCC How-to Guides. Edinburgh: Digital Curation Centre. 2010. URL: dcc.ac.uk/resources/how-guides/appraise-select-data.

11. **Data** management costing tool and checklist, version 3, UK Data Service. UK Data Service (UKDS). 2015. URL: ukdataservice.ac.uk/manage-data/plan/costing.
12. **DCC**. Five steps to decide what data to keep: checklist for appraising research data // Edinburgh: Digital Curation Centre. 2014. V. 1. URL: dcc.ac.uk/resources/how-guides.
13. **Tenopir C., Rice N. M., Allard S., Baird L., Borycz J., Christian L. et al.** Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide // PLoS ONE. 2020. V. 15. № 3. e0229003. DOI 10.1371/journal.pone.0229003.
14. **Beagrie N.** What to Keep: A Jisc research data study. 2019. URL: <https://apo.org.au/node/227001>.
15. **Zalaev G. Z., Tcvetkova V. A.** Nekotory'e voprosy' khraneniia tcifrovoi' informatcii // NTI-2022. Nauchnaia informatciia v sovremennom mire: global'ny'e vy'zovy' i natsional'ny'e priorityty' : materialy' 10-i' nauchnoi' konferencii s mezhdunarodny'm uchastiem, posviashchyonnoi' 70-letiiu VINITI RAN, Moskva, 25–26 oktiabria 2022 g. Moskva : Vserossii'skii' institut nauchnoi' i tekhnicheskoi' informatcii RAN, 2022. S. 73–75. DOI 10.36535/2022-9785945770829-14.
16. **Zalaev G. Z., Kalenov N. E., Tcvetkova V. A.** Otcifrovka dokumentov v nauchny'kh arhivakh i bibliotekakh: voprosy' i otvety' // NTI. Ser. 1: Organizatciia i metodika informatcionnoi' raboty'. 2016. № 2. S. 14–21.
17. **Soohyung J., Hofman D., Youngseek K.** Investigation of challenges in academic institutional repositories: A survey of academic librarians // Library Hi Tech. 2019. V. 37. № 3. Pp. 525–548.
18. **Tait E.** Libraries for the future: the role of IT utilities in the transformation of academic libraries // Palgrave Communications. 2016. V. 2. 16070. DOI 10.1057/palcomms.2016.70.
19. **Roche D. G., Lanfear R., Binning S. A., Haff T. M., Schwanz L. E., Cain K. E.** Troubleshooting Public Data Archiving: Suggestions to Increase Participation // PLoS Biology. 2014. V. 12. № 1. e1001779. DOI 10.1371/journal.pbio.1001779.
20. **Salomon J. A.** Stanford University School of Medicine, Stanford, CA, and accepted by the Editorial Board October 18 // PNAS. 2021. V. 118. e2111454118.
21. **Xueqin Guo, Fengzhen Chen, Fei Gao, Ling Li, Ke Liu, Lijin You, Cong Hua, Fan Yang, Wanliang Liu, Chunhua Peng, Lina Wang, Xiaoxia Yang, Feiyu Zhou, Jiawei Tong, Jia Cai, Zhiyong Li, Bo Wan, Lei Zhang, Tao Yang, Minwen Zhang, Linlin Yang, Yawen Yang, Wenjun Zeng, Bo Wang, Xiaofeng Wei, Xun Xu.** CNSA: a data repository for archiving omics data, Database. 2020, 2020:baaa055. DOI 10.1093/database/baaa055.
22. **Fevola C., Görgen C.** The mathematical research-data repository MathRepo // arXiv preprint arXiv:2202.04022. 2022. URL: <https://arxiv.org/abs/2202.04022>.
23. **Rozhdstvenskaia M. Iu.** Repozitorii' kak realizatciia idei' otkry'togo dostupa k nauchny'm publikatciam: podhody' k klassifikatcii // Bibliosfera. № 2. 2015. Pp. 86–94.
24. **Kouper I.** Professional participation in digital curation // Library & Information Science Research. 2016. T. 38. № 3. C. 212–223.
25. **Kowalczyk S. T.** Digital curation for libraries and archives // Bloomsbury Publishing USA. 2018. ISBN 978-1610696319.

26. **Jørn Nielsen H., Hjørland B.** Curating research data: the potential roles of libraries and information professionals // *Journal of documentation*. 2014. T. 70. № 2. С. 221–240.
27. **Tiwari Amit.** Data analytics in libraries and information centres. 2018. URL: https://www.researchgate.net/publication/327389981_data_analytics_in_libraries_and_information_centres.
28. **Tercan H., Meisen T.** Machine learning and deep learning based predictive quality in manufacturing: a systematic review // *Journal of Intelligent Manufacturing*. 2022. V. 33. Pp. 1879–1905 (2022). DOI 10.1007/s10845-022-01963-8.
29. **Biswas A.** Application of Data Analytics for Mapping of Library System and Services of Oxford University Library. 2023. URL: https://www.researchgate.net/publication/375447227_Application_of_Data_Analytics_for_Mapping_of_Library_System_and_Services_of_Oxford_University_Library.
30. **Hsin-liang Chen, Philip Doty, Carol Mollman, Xi Niu, Jen-chien Yu, Tao Zhang.** Library assessment and data analytics in the big data era: Practice and policies // *Proceedings of the Association for Information Science and Technology*. 2020. V. 52. I. 1. Pp. 1–4. DOI 10.1002/pr2.2015.14505201002.
31. **Preza Díaz J. L.** Data Science and Analytics in Libraries // *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*. 2017. V. 70. P. 244. DOI 10.31263/voebm.v70i2.1796.
32. **Red'kina N. S.** Vektory` razvitiia nauchny`kh bibliotek: obzor cluichevy`kh docladov Vsemirnogo kongressa IFLA 2019 g. // *Bibliosfera*. 2020. № 2. S. 71–81. DOI 10.20913/1815-3186-2020-2-71-81.
33. **Kvetkina E. A.** Leechny`i` kabinet pol`zovatel'ia kak virtual`ny`i` servis informatcionnogo soprovozhdeniia biblioteko`i` nauchno-issledovatel'skoi` deiatel`nosti. Repozitorii` TCNB NAN Belarusi. 2023. S. 136–143.
34. **Gorbich L. G.** Informatcionnaia sistema «Web-kabinet uchyonogo» kak interaktivnaia e`lektronnaia biblioteka // *Biblioteki vuzov Urala: problemy` i opyt` raboty`*. 2015. V. 14. № 14. S. 70–72.
35. **Mitroshin I. A.** Osnovny`e printipy` razvitiia sai`ta nauchnoi` biblioteki // *Nauchny`e i tekhnicheskie biblioteki*. 2020. № 11. S. 165–184. DOI 10.33186/1027-3689-2020-11-165-184.
36. **Nuriev V. A., Egorova A. Iu.** Metody` ocenki kachestva mashinnogo perevoda: sovremennoe sostoianie, Informatika i eyo primenenie. 2021. T. 15. V. 2. S. 104–111. DOI 10.14357/19922264210215.
37. **Konev K. A.** Ispol`zovanie metodov mashinnogo obucheniia v zadachakh priiniatiia reshenii` pri obespechenii kachestva v priborostroenii // *E`konomika. Informatika*. 2022. V. 49. № 4. Pp. 820–832.
38. **Veretennikov I. S., Kartashev E. A., Tcaregorodtcev A. L.** Ocenka kachestva classifikatcii tekstovy`kh materialov s ispol`zovaniem algoritma mashinnogo obucheniia «Sluchai`ny`i` les» // *Izvestiia Altai`skogo gosudarstvennogo universiteta*. 2017. № 4 (96). Pp. 78–83.
39. **Kleijnen Jack & van Groenendaal Willem.** Measuring The Quality Of Publications: New Methodology And Case Study // *Inf. Process. Manage*. 2000. V. 36. Pp. 551–570. DOI 10.2139/ssrn.247676.

40. **Mike Thelwall.** Can the quality of published academic journal articles be assessed with machine learning? *Quantitative Science Studies*. 2022. V. 3. № 1. Pp. 208–226. DOI 10.1162/qss_a_00185.
41. **Zhou J., Gandomi A. H., Chen F., Holzinger A.** Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics // *Electronics*. 2021. V. 10. P. 593. DOI 10.3390/electronics10050593.
42. **Becker C., Kulovits H., Guttenbrunner M. et al.** Systematic planning for digital preservation: evaluating potential strategies and building preservation plans // *International Journal on Digital Libraries*. 2009. V.10. Pp. 133–157 DOI:10.1007/s00799-009-0057-1.
43. **Van der Hoeven J., van Wijngaarden H.** Modular emulation as a long-term preservation strategy for digital objects // In: 5th International Web Archiving Workshop (IWAW05). Vienna, Austria. 2005. URL: <https://api.semanticscholar.org/CorpusID:9066566>.
44. **Begishev I. R.** Pravovy'e sredstva obespecheniia bezopasnosti tcifrovoy'kh arhivov v usloviiahk vnedreniia tekhnologii' iskusstvennogo intellekta // *Vestnyk luridicheskogo instituta MIIT*. 2021. № 2 (34). S. 108–116.
45. **El Kafhali S., El Mir I., Hanini M.** Security Threats, Defense Mechanisms, Challenges, and Future Directions in Cloud Computing // *Archives of Computational Methods in Engineering*. 2022. V. 29. 223–246. DOI 10.1007/s11831-021-09573-y.
46. **Alcaraz C., Lopez J.** Digital Twin: A Comprehensive Survey of Security Threats // *IEEE Communications Surveys & Tutorials*. 2022. Vol. 24. DOI 10.1109/COMST.2022.3171465.
47. **Yeh P. S., Moury G. A., Armbruster P.** CCSDS data compression recommendations: development and status // *Applications of Digital Image Processing XXV* // *SPIE*. 2002. V. 4790. Pp. 302–313.
48. **Zierau E., Schultz M.** Creating a framework for applying OAIS to distributed digital preservation // *Proceedings of the 10th International Conference on Preservation of Digital Objects*. 2013. Pp. 78–83.
49. **Hodge G.** Archiving and Preservation in Electronic Libraries // *Electronic Information Management for PFP Nations*. 2002. URL: https://www.researchgate.net/publication/235146965_Archiving_and_Preservation_in_Electronic_Libraries.
50. **Ledoux T.** Long-term preservation at the National Library of France (BnF): Scalable Preservation and Archiving Repository (SPAR) // *International Preservation*. 2012. Pp. 2. URL: https://irods.org/uploads/2014/06/BnF_2014_06_19_iRodsUG_BnF-ADM-2014-061850-01-p2.pdf.
51. **Lounamaa K.** EVA – The Acquisition and Archiving of Electronic Network Publications In Finland. 1999. URL: <https://www.ercim.eu/publication/ws-proceedings/DELOS6/eva.pdf>.
52. **Al Noamany Y., Borghi J. A.** Towards computational reproducibility: researcher perspectives on the use and sharing of software // *PeerJ Computer Science*. 2018. № 4. e163 DOI 10.7717/peerj-cs.163.
53. **Housewright R., Schonfeld R. C., Wulfson K.** US faculty survey 2012 // *Ithaka S + R*. 2013. Pp. 45–80. DOI 10.18665/sr.22502.

Информация об авторах / Authors

Бескаравайная Елена Вячеславовна –

старший научный сотрудник
Библиотеки по естественным
наукам РАН, Москва, Российская
Федерация

elenabesk@gmail.com

Митрошин Иван Андреевич –

старший научный сотрудник
Библиотеки по естественным
наукам РАН, Москва, Российская
Федерация

imitros@gmail.com

Elena V. Beskaravainaya – Senior

Researcher, Library for Natural
Sciences, Russian Academy
of Sciences, Moscow, Russian
Federation

elenabesk@gmail.com

Ivan A. Mitroshin – Senior Researcher,

Library for Natural Sciences,
Russian Academy of Sciences,
Moscow, Russian Federation

imitros@gmail.com