

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА  
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 11

Москва 2025

## ОБЩИЙ РАЗДЕЛ

УДК [004.8:001]:165

А.Н. Тимохович, Е.В. Самоходкин, А.А. Эльзон

### Классификация эпистемологических угроз искусственного интеллекта

*Рассматриваются систематизация и классификация эпистемологических угроз, возникающих в результате интеграции искусственного интеллекта в научную деятельность. На основе анализа эмпирических и теоретических данных выделены ключевые категории рисков: утрата достоверности научного знания, когнитивные искажения, снижение научного плюрализма, а также эрозия доверия и этических оснований научного сообщества. Обоснована необходимость комплексного подхода к оценке и минимизации выявленных угроз для обеспечения устойчивости современной науки.*

**Ключевые слова:** искусственный интеллект, эпистемология, научное знание, генеративные модели, ИИ-угрозы, научный плюрализм, доверие

DOI: 10.36535/0548-0019-2025-11-1

#### ВВЕДЕНИЕ

Эмпирические данные последних лет демонстрируют экспоненциальный рост внедрения алгоритмов искусственного интеллекта в сферу академической

науки. Согласно аналитическим материалам Pew Research Center и FastDataScience, уже к 2020 г. свыше 50% исследовательских проектов в ведущих научных институтах опирались на технологии искус-

ственного интеллекта для обработки данных, построения моделей и автоматизации отдельных этапов научного процесса, тогда как в 2010 г. этот показатель составлял менее 20%. Наблюдается существенное увеличение доли публикаций, тематически связанных с искусственным интеллектом: с 5% от общего числа научных статей в 2010 г. до 30% в 2020-м г. [1]. Эти данные подтверждают не только интеграцию искусственного интеллекта в науку, но трансформацию самой логики научного процесса.

Согласно крупномасштабному библиометрическому анализу 67,9 млн научных публикаций, ученые, использующие инструменты искусственного интеллекта в своей исследовательской практике, демонстрируют более высокую продуктивность: в среднем объем их публикационной активности выше на 67,37%, а индекс цитируемости превышает контрольную выборку в 3,16 раза [2]. Однако выявлен и значимый эпистемологический сдвиг, связанный с использованием генеративных моделей искусственного интеллекта, таких как большие языковые модели (LLMs – Large Language Models), которые могут искажать или противоречить традиционным методам научного познания.

Одна из ключевых проблем – это феномен «коллапса знания», при котором чрезмерная зависимость от технологий искусственного интеллекта приводит к снижению разнообразия и глубины научных исследований. Ряд ученых отмечает, что широкое использование генеративных моделей может привести к унификации научных подходов, ограничивая инновации и критическое мышление [3–5]. Непрозрачность алгоритмов глубокого обучения создает препятствия для понимания и интерпретации научных результатов. Так, E. Duede подчеркивает, что использование «черных ящиков» в научных исследованиях затрудняет обоснование и воспроизводимость результатов, что является фундаментальным принципом научного метода [6].

Этические и интегральные риски также становятся все более актуальными. H. Balalle и S. Pannilage, проведя систематический обзор литературы в контексте оценки влияния искусственного интеллекта на академическую добросовестность, указывают на необходимость пересмотра этических норм и процедур в свете новых вызовов, связанных с использованием искусственного интеллекта в науке, включая вопросы авторства, ответственности и достоверности данных [7]. В свою очередь, E. Papagiannidis, P. Mikalef и K. Conboу предлагают концептуальную основу для ответственного управления искусственным интеллектом, включающую принципы подотчетности, разнообразия, недискриминации и справедливости, человеческой субъектности (human agency – способность человека к сознательному, автономному действию и принятию решений, основанных на собственных мотивах, целях и убеждениях) и т.д. [8]. Дополнительно авторы отмечают необходимость разработки и внедрения этических принципов и руководств для обеспечения ответственного использования искусственного интеллекта в различных сферах, включая науку.

Настоящее исследование направлено на систематизацию и классификацию эпистемологических угроз, возникающих в результате интеграции искусственного интеллекта в научную деятельность. Объект исследования – эпистемологические угрозы, обусловленные развитием и внедрением систем искусственного интеллекта, предметом следует считать классификацию эпистемологических угроз, порождаемых применением искусственного интеллекта, а цель заключается в разработке классификации эпистемологических угроз, формируемых последствиями использования искусственного интеллекта в сфере науки.

Отметим, что существенным ограничивающим фактором настоящей работы выступает характер современных исследований в области эпистемологических рисков применения искусственного интеллекта: большинство научных разработок, опубликованных в период 2020–2025 гг., как наиболее содержательно насыщенный с точки зрения радикальных трансформаций в области разработки и применения искусственного интеллекта, носят преимущественно теоретико-обзорный характер и лишь в редких случаях опираются на масштабные эмпирические или количественные методы. Это затрудняет верификацию отдельных положений, а также ограничивает возможности для построения строгих причинно-следственных моделей влияния искусственного интеллекта на научное знание.

Соответственно, интеграция искусственного интеллекта в сферу академической науки сопряжена не только с трансформацией научных практик и изменением структуры исследовательской деятельности, но и с возникновением новых классов эпистемологических угроз, требующих систематического анализа. Накапливающиеся эмпирические и теоретические данные акцентируют необходимость перехода от разрозненных наблюдений к комплексной методологии, позволяющей формализовать критерии выявления и классификации рисков, обусловленных внедрением искусственного интеллекта. Сложность и многогранность возникающих вызовов, включая феномен коллапса знания, утрату научного плюрализма, непрозрачность алгоритмических решений, а также угрозы академической этике и доверию, обуславливают актуальность поиска инструментов количественной и качественной оценки анализируемых вызовов.

Представляется, что в современных условиях наиболее эффективна разработка методологических подходов, способных интегрировать сравнительный, формально-логический и эмпирико-статистический анализ для диагностики и прогностической верификации эпистемологических угроз, формируемых искусственным интеллектом в сфере научного знания. Такая стратегия не только создает предпосылки для выработки механизмов управления и минимизации негативных последствий применения технологий искусственного интеллекта, но и содействует укреплению устойчивости научного знания в эпоху радикальных технологических преобразований. Результаты настоящего исследования вносят вклад в построение целостной концепции оценки и классификации рисков, вызванных развитием искусственного

интеллекта, что, в свою очередь, имеет большое значение для обеспечения безопасности, адаптивности и устойчивости современной научной среды.

## ОСНОВАНИЯ И МЕТОДИКА КЛАССИФИКАЦИИ

Развитие искусственного интеллекта ставит перед наукой новые эпистемологические риски – угрозы, способные нарушать течение процессов возникновения, проверки и распространения научного знания. Для систематического анализа рисков необходимо понимание их типов. Классификация должна опираться на критерии, отражающие природу угроз, а именно: механизм возникновения, стадия научного познания, степень универсальности моделей искусственного интеллекта.

Во-первых, угрозы различаются по механизму их возникновения: одни угрозы вытекают из технических ограничений и свойств искусственного интеллекта (например, ошибки генеративных моделей; высокий bias данных, свидетельствующий о недостаточной гибкости модели; «черный ящик» алгоритмов), другие – из когнитивного взаимодействия искусственного интеллекта и ученых (например, избыточное доверие к результатам или существующие искажения восприятия), третьи – из социально-инфраструктурных факторов (отсутствие или низкий уровень рецензирования, появление псевдонаучных публикаций, снижение доверия к экспертам).

Во-вторых, критерий стадий научного познания позволяет разграничить угрозы, проявляющиеся на разных этапах существования научного знания: на этапе формирования научного знания (формулировка гипотез, сбор и анализ данных); на этапе валидации (проверка, рецензирование, воспроизводимость); на этапе диссеминации (публикация результатов научных исследований, обсуждение в научном сообществе).

В-третьих, следует учитывать различия между узкоспециализированными моделями искусственного интеллекта и универсальными (широкопрофильными) моделями. Алгоритмы с ограниченной функциональной направленностью, обученные для решения конкретных задач, могут вносить систематические ошибки в пределах доменной области (например, bias в медицинских изображениях или физических экспериментах), тогда как крупные интегральные модели общего назначения (такие как LLMs, включая GPT, PaLM, Claude и т.п.) способны генерировать тексты и выводы в любой области знания, что расширяет масштаб и непредсказуемость возникающих рисков.

Таким образом, выделенные критерии – механизм возникновения, стадия научного познания, степень универсальности моделей искусственного интеллекта – обеспечивают многомерную сетку для разработки классификации эпистемологических угроз. Далее будет представлена развернутая классификация эпистемологических угроз искусственного интеллекта (ИИ-угроз), в рамках которой акцентируется влияние на достоверность научных результатов, а также на социальные механизмы функционирования науки.

## ОСНОВНЫЕ ТИПЫ ЭПИСТЕМОЛОГИЧЕСКИХ ИИ-УГРОЗ

Первостепенно выделим угрозы достоверности и надежности научного знания. Мощные системы искусственного интеллекта способны генерировать привлекательные, но недостоверные результаты. Базовые модели (foundation models), например, языковые модели LLM, склонны к так называемым «галлюцинациям» (hallucinations) – формулировке несуществующих фактов, ссылок и выводов. Например, в отчете OpenAI [9, с. 4] эта проблема рассматривается в контексте новых языковых моделей o3 и o4-mini. Представлены количественные оценки, демонстрирующие, что модели o3 и o4-mini имеют более высокие показатели галлюцинаций по сравнению с предыдущими версиями. В частности, на бенчмарке PersonQA, предназначенном для оценки точности ответов моделей на вопросы о публичных личностях, модель o3 продемонстрировала уровень галлюцинаций в 33%, что примерно вдвое превышает показатели предыдущих моделей o1 (16%) и o3-mini (14,8%). Модель o4-mini показала еще более высокий уровень галлюцинаций – 48%. Эти данные указывают на то, что, несмотря на улучшение в ряде других характеристик, новые модели демонстрируют более высокую склонность к генерации недостоверной информации, что подчеркивает парадоксальность соотношения между архитектурным прогрессом и эпистемической надежностью вывода.

Например, нейросеть может уверенно предоставить вымышленную научную ссылку или неверный вывод, который выглядит правдоподобно, и тем самым ввести исследователей в заблуждение [10]. Даже узкоспециализированные модели искусственного интеллекта могут совершать скрытые ошибки: их выводы основаны на корреляциях в данных, а не на причинных связях, что создает риск принятия ложных гипотез [11]. Особую обеспокоенность вызывает систематический bias (систематическое отклонение от объективности) и искажение данных: искусственный интеллект не только наследует предвзятость обучающих выборок, но порой и усиливает их. Исследование UNESCO 2024 г. выявило, что LLMs значительно чаще ассоциируют женщин с домашними ролями и семьей, тогда как мужчин – с карьерой и лидерством (в одной модели женщины описывались в бытовых ролях в четыре раза чаще) [11]. Такие примеры демонстрируют, как алгоритмы воспроизводят гендерные и другие стереотипы, угрожая объективности научного знания.

Непрозрачность (неинтерпретируемость) алгоритмов осложняет выявление и исправление ошибок. Когда модель является так называемым «черным ящиком» (black box – модель, чьи решения невозможно логически реконструировать на уровне причинно-следственных связей, доступных для внешнего наблюдателя), ученым трудно понять, на чем основан ее вывод; это подрывает классические эпистемические критерии воспроизводимости и проверяемости. Кризис воспроизводимости науки усугубляется искусственным интеллектом: по данным аналитического обзора Организации экономического

сотрудничества и развития (ОЭСР), проблема невозможности воспроизводимых результатов затрагивает большинство дисциплин, и использование сложных моделей искусственного интеллекта становится одним из факторов невозможности воспроизведения результатов научного исследования при повторении эксперимента или анализа по тем же методам и данным, что и в исходной работе, т.е. формируя кризис воспроизводимости (reproducibility crisis) [12, с. 262]. Таким образом, первая категория угроз связана с возможной утратой достоверности научных результатов из-за ошибок, галлюцинаций и bias, привнесенных искусственным интеллектом.

Во-вторых, важно выделить угрозы пониманию и объяснению в силу когнитивных иллюзий и чрезмерного доверия. Внедрение искусственного интеллекта в научную практику меняет эпистемологию самого понимания. С одной стороны, искусственный интеллект обещает преодолеть человеческие ограничения – повысить продуктивность, обеспечить объективность анализа. С другой стороны, эти обещания могут эксплуатировать когнитивные ограничения человека, порождая иллюзии понимания. Ученые, впечатленные «умными» алгоритмами, нередко переоценивают степень понимания, которую предоставляет им ИИ-инструмент. В исследовании L. Messeri и M.J. Crockett отмечается, что увлечение искусственным интеллектом грозит наступлением фазы науки, в которой мы производим больше, но понимаем меньше [13]. Авторы выделяют целый ряд таких иллюзий: во-первых, иллюзия объяснительной глубины, когда наличие точного прогноза от модели внушает исследователю ощущение, будто он выяснил причину возникновения явления, хотя на самом деле понимание остается поверхностным; во-вторых, иллюзия широты обзора, когда благодаря искусственному интеллекту у ученого создается впечатление, будто он проверил все выдвинутые гипотезы, хотя алгоритм мог игнорировать альтернативные подходы; в-третьих, иллюзия объективности, когда ИИ-инструмент ошибочно воспринимается как беспристрастный, как эксперт высокого уровня, которому можно полностью доверять.

Такие когнитивные ловушки основаны на авторитете машинной точности. Возникает феномен повышенного доверия к искусственному интеллекту: человеку свойственно проецировать на алгоритм черты рационального агента и доверять выводам машины даже больше, чем собственному суждению [14]. Например, в работе С. Holbrook и коллег были проведены два эксперимента, направленные на изучение склонности людей к повышенной доверчивости к рекомендациям искусственного интеллекта в контексте принятия решений об идентификации образа врага [15]. Участникам эксперимента предлагалось принять участие в симулированной задаче по идентификации угроз, где они должны были классифицировать визуальные стимулы как представляющие врагов или гражданских лиц. После первоначальной классификации участники получали обратную связь от ИИ-агента, который случайным образом соглашался или не соглашался с их решением. Затем участники могли изменить свое первоначальное решение и оценить свою уверенность в нем. При расхождении оценок

участников эксперимента и искусственного интеллекта 58,3% респондентов в первом эксперименте и 67,3% во втором изменяли свое решение в соответствии с рекомендацией искусственного интеллекта. Первоначальная точность классификации участников составляла 72,1% в первом эксперименте и 65,0% во втором. Однако при несовпадении оценок искусственного интеллекта точность снижалась до 53,8% и 41,3% соответственно. Участники, чьи решения совпадали с рекомендациями искусственного интеллекта, сообщали о повышении уверенности в своих решениях. Напротив, при несоответствии оценок искусственного интеллекта и сохранении первоначального решения, уверенность участников снижалась. Полученные результаты исследования демонстрируют значительную склонность людей к повышенной доверчивости к рекомендациям искусственного интеллекта, даже когда эти рекомендации являются случайными и потенциально ошибочными.

В рамках эмпирического исследования, проведенного D. Ullrich, A. Butz и S. Diefenbach, была предпринята попытка детально проанализировать феномен избыточного доверия к автономным роботизированным системам в контексте бытового сценария, построенного на взаимодействии с автоматизированной кормушкой для домашних животных [16]. В качестве основы использовалась симулированная ситуация, в которой испытуемые, численностью 110 человек, воображали себя участниками 28-дневного сафари, оставляя домашнего кота под присмотром автономного устройства. Каждый день в течение симуляции участникам необходимо было принимать решение относительно собственной стратегии поведения: либо довериться роботу и продолжить путешествие, либо предпринять дополнительные действия для проверки его работы и, при необходимости, задействовать резервную процедуру аварийной активации. Такие экспериментальные условия позволили смоделировать типичный для взаимодействия с ИИ-диспозитивами выбор между доверием и контролем, между автоматическим делегированием ответственности и когнитивной вовлеченностью пользователя.

В течение первых трех недель робот работал без сбоев, что способствовало росту доверия участников. Однако на четвертой неделе робот трижды подряд не выполнил свою задачу, что могло привести к гибели кота, если бы участники не осуществляли проверку.

Дополнительно исследование рассматривало влияние на уровень доверия трёх факторов: личного опыта взаимодействия с роботом, репутации, основанной на отзывах других пользователей, и демонстрации возможностей через видеоматериалы. Результаты показали, что личный опыт оказывал наиболее значительное влияние на формирование доверия, в то время как репутация и демонстрации имели лишь временный эффект. Со временем по мере накопления положительного опыта взаимодействия с роботом, уровень доверия участников возрастал, и количество контрольных проверок снижалось. В результате, в последней фазе симуляции, когда робот начал давать сбои, многие участники продолжали полагаться на него без дополнительной проверки, что приводило к негативным последствиям.

Итак, вторая категория угроз – эпистемические искажения на уровне понимания – включает в себя опасность утраты критического мышления, чрезмерного доверия к автоматизированным «соавторам» научного процесса. Когнитивные bias, такие как автоматизм доверия (automation bias), могут привести к тому, что исследователь не перепроверяет вывод, если его предложил искусственный интеллект, что подрывает принцип скептицизма, фундаментальный для научного метода.

В-третьих, необходимо отметить угрозу разнообразию и инновационности научного поиска. Научное знание устойчиво, когда оно опирается на множество конкурирующих гипотез, методов и точек зрения [17]. Масштабное внедрение одинаковых ИИ-инструментов, особенно LLMs от ограниченного числа разработчиков, несет риск формирования научной монокультуры [13]. Если все лаборатории будут использовать сходные алгоритмы для анализа данных и генерации гипотез, есть опасность, что одни и те же типы вопросов и подходов начнут доминировать, вытесняя альтернативные направления исследования.

Например, A.A.H. de Hond и коллеги рассматривают влияние демографической однородности команд разработчиков искусственного интеллекта на разнообразие создаваемых решений [18]. Результаты исследования показывают, что команды с низким уровнем разнообразия склонны разрабатывать модели с ограниченным спектром решений, а это может приводить к системным ошибкам и снижению эффективности алгоритмов. В свою очередь, J. Volle и коллеги применяют количественно-аналитический подход (иерархические многослойные байесовские стохастические блочные модели для изучения сетей сотрудничества, цитирования и тематических направлений) для моделирования процессов формирования научных групп в США и Германии с целью выявления динамики гомогенизации и дифференциации в научных сообществах этих стран в период 2004–2019 гг. [19]. Установлено, что в американской социологии наблюдается тенденция к гомогенизации, выражающаяся в формировании крупных научных групп с высокой степенью внутренней связанности. Это может приводить к снижению разнообразия исследовательских подходов и тем. В немецкой социологии, напротив, прослеживается тенденция к дифференциации с формированием множества небольших и специализированных научных групп, что способствует поддержанию разнообразия исследовательских направлений. Различия в институциональных структурах и научной политике между США и Германией оказывают значительное влияние на процессы формирования научных групп и, соответственно, на уровни гомогенизации и дифференциации в этих группах. Исследование подчеркивает, что, с одной стороны, гомогенизация может способствовать усилению коллабораций и концентрации ресурсов, с другой – дифференциация обеспечивает разнообразие подходов и инновационность. Авторы [19] призывают к осознанной научной политике, учитывающей эти аспекты для устойчивого развития науки.

Следует подчеркнуть, что искусственный интеллект, обучаемый преимущественно на исторических

массивах научных данных, по своей сути склонен к экстраполяции уже сложившихся эмпирических и теоретических паттернов. Такая архитектурная особенность способствует институционализации и, в определенной степени, консервации существующего научного статус-кво. В результате, вместо активизации возможных революционных сдвигов и стимулирования парадигматических трансформаций, о которых упоминал Т.С. Кун, возможно дальнейшее укрепление господствующих исследовательских программ, некритически рекомендуемых алгоритмическими системами как наиболее вероятные или валидные [20]. Тем самым возрастает риск замедления или даже блокировки появления новых концептуальных подходов, что ведет к гомогенизации научного дискурса и снижению эпистемологического многообразия.

Более того, существующие социальные систематические отклонения от объективности могут транслироваться искусственным интеллектом в выборе тем исследований: например, если данные смещены в пользу изучения заболеваний в развитых странах, то алгоритмы будут реже предлагать исследования по проблемам развивающихся стран, усугубляя эпистемическое неравенство. Исходя из этого, важно отметить, что особого внимания заслуживает проблема переноса и обобщения моделей, обученных на данных из высокоразвитых стран, на условия с ограниченными ресурсами.

Так, исследование международной исследовательской группы под руководством С. Sendra-Balcells было направлено на оценку способности моделей глубокого обучения, разработанных для классификации стандартных плоскостей фетального ультразвукового исследования, к обобщению на данные, полученные в условиях с ограниченными ресурсами [21]. Первоначально модель была обучена на данных 1792 пациентов из Испании, после чего ее эффективность протестирована на выборке из 1008 пациентов в Дании, где условия сбора данных аналогичны исходным. Затем модель была адаптирована для применения в пяти африканских странах: Алжире, Гане, Египте, Малави и Уганде, каждая из которых предоставила по 25 пациентов для тестирования. Результаты показали, что без дополнительной адаптации производительность модели значительно снижалась при применении к данным из африканских стран, что свидетельствует о наличии так называемого «смещения домена» (domain shift). Однако использование подхода трансферного обучения позволило повысить полноту (recall) модели до  $0.92 \pm 0.04$ , при этом сохранив высокую точность (precision) классификации. Полученные результаты подчеркивают важность учета контекста и специфики данных при разработке и применении ИИ-моделей в различных условиях. Игнорирование этих факторов может привести к усилению существующих диспропорций и ограничению доступности качественной медицинской помощи в странах с ограниченными ресурсами.

Работа L.A.Celi и коллег [22] направлена на выявление и систематизацию источников предвзятости в клинических ИИ-системах, которые могут способствовать усилению существующих неравенств в здравоохранении. Авторы сосредоточились на анали-

зе происхождения обучающих данных, демографических характеристик исследовательских коллективов и распределения клинических специальностей в публикациях, посвященных искусственному интеллекту в медицине. Один из ключевых выводов работы – это значительная географическая концентрация данных: более 50% используемых для обучения клинических ИИ-моделей датасетов происходят из США и Китая. Свыше 40% авторов соответствующих научных публикаций также представляют эти две страны. Такое распределение указывает на ограниченное представительство других регионов, особенно стран с низким и средним уровнем доходов, что может привести к снижению эффективности ИИ-моделей при их применении в этих странах. Дополнительно, анализ авторского состава выявил половой и профессиональный дисбаланс: почти 75% позиций первого и старшего автора занимают мужчины, а около 40% всех публикаций сосредоточено в области радиологии. Такое распределение может ограничивать разнообразие взглядов и подходов в разработке ИИ-систем, способствуя усилению существующих предвзятостей. Таким образом, это исследование подчеркивает необходимость более широкого представительства различных регионов, полов и профессиональных специальностей в разработке и применении ИИ в медицине.

Философы науки вводят концепцию «эпистемической несправедливости» (epistemic injustice), когда отдельные группы систематически исключаются из процесса производства знаний. Например, J. Kay, A. Kasirzadeh и S. Mohamed анализируют, каким образом генеративные ИИ-системы могут подрывать целостность коллективного знания и процессы, на которые научное сообщество полагается для получения, оценки и доверия к информации, представляя собой значительную угрозу для информационной экосистемы и демократического дискурса [23]. Это исследование основано на междисциплинарном подходе, объединяющем социальную и политическую философию, а также на анализе конкретных примеров использования генеративных ИИ-систем. Авторы вводят концепцию «генеративной алгоритмической эпистемической несправедливости» и выделяют четыре ключевых измерения этого феномена: усиленная свидетельская несправедливость, манипулятивная свидетельская несправедливость, герменевтическое неведение и герменевтическая несправедливость доступа.

1. Усиленная свидетельская несправедливость проявляется в том, что генеративные модели, обученные на данных, содержащих социальные предвзятости, могут усиливать эти предвзятости в своих ответах. Например, при генерации текстов или изображений, описывающих определенные профессии или социальные роли, модели могут непропорционально представлять различные демографические группы, что приводит к искаженному восприятию реальности.

2. Манипулятивная свидетельская несправедливость возникает, когда пользователи сознательно используют генеративные ИИ-системы для создания и распространения дезинформации. Примером может служить использование генеративных моделей для создания фальшивых изображений или текстов, направленных на дискредитацию определенных групп или

распространение ложных сведений в политических целях.

3. Герменевтическое неведение связано с неспособностью генеративных моделей адекватно интерпретировать и представлять опыт маргинализированных групп из-за отсутствия соответствующих данных в обучающих выборках. Это приводит к тому, что определенные культурные или социальные контексты оказываются недоступными или искаженными в ответах ИИ-систем.

4. Герменевтическая несправедливость доступа проявляется в ограниченном доступе к знаниям и информации для определенных групп, особенно в многоязычных контекстах. Генеративные ИИ-системы, преимущественно обученные на данных на английском языке, могут предоставлять менее точные или неполные ответы на других языках, что усугубляет существующие информационные неравенства.

Авторы [23] подчеркивают, что генеративные ИИ-системы, обученные на больших объемах данных, часто отражают предвзятости, присущие этим данным, что может приводить к систематическому искажению информации и усилению существующих социальных и культурных неравенств. Особое внимание уделяется тому, как такие системы могут неадекватно представлять или полностью игнорировать знания и опыт маргинализированных групп. Авторы призывают к созданию информационной экосистемы, которая поддерживает целостность производства знаний.

На основе приведенного анализа литературных источников представляется возможным выделить особую категорию эпистемологических рисков, связанных с угрозой утраты научного плюрализма в результате доминирования генеративных моделей, обученных преимущественно на англоязычных западных корпусах данных. Такая предрасположенность к воспроизведению гегемониальных нарративов неизбежно ведет к маргинализации локальных языков, традиций и иных уникальных проблемных полей, сужая тем самым глобальную повестку исследовательских инициатив и ослабляя мозаичность научного знания. В подобном контексте третья группа угроз может быть концептуализирована как риск стагнации научной мысли и эрозии многообразия перспектив, когда интеллектуальный процесс фиксируется в границах «рекомендованного» ИИ научного канона. Без должных условий для поддержания эпистемического разнообразия и критического противостояния альтернативных идей существует опасность того, что наука утратит способность к радикальным трансформациям, инновациям и пересмотру устоявшихся доктрин. Следует отметить, что рассматриваемый класс рисков присущ как специализированным, так и универсальным архитектурам искусственного интеллекта. Алгоритмы с узкой предметной областью, будучи оптимизированы на типовых паттернах, могут автоматически отбрасывать «аномальные» или нетривиальные данные как статистический шум, тем самым не способствуя появлению новых научных направлений. В свою очередь, ИИ-модели общего назначения, характеризующиеся тенденцией к усреднению и репродукции наиболее частотных и популярных теорий, способны укреплять эпистемическую гомогени-

зацию за счет подавления инновационных и маргинальных концепций.

Заключительной, четвертой угрозой предлагается считать угрозу инфраструктуре доверия и интегритету (integrity – целостность, честность и соблюдение фундаментальных принципов добросовестности в научной практике) научного сообщества. Помимо качества самих знаний, искусственный интеллект затрагивает социальные механизмы, через которые научное знание утверждается и распространяется. Современная научная инфраструктура – рецензирование, издание журналов, библиометрия – испытывает беспрецедентное давление [24]. С одной стороны, искусственный интеллект способен облегчить создание научной работы: уже появляются инструменты для автоматического обзора литературы и написания черновиков статей, такие как нейропоиск от eLibrary, информационная система научной аналитики «НАУКА-МГУ», Elicit, Scite.ai, Research Rabbit и др. С другой стороны, такая автоматизация может нарушить научный процесс. Например, автоматизированные обзоры литературы, генерируемые за секунды, грозят снижением качества осмысления результатов предыдущих исследований. Отмечается, что чрезмерная скорость в подготовке обзоров с помощью искусственного интеллекта может подрывать непосредственно прогресс науки, поскольку ученые рискуют пропустить противоречия и тонкости, которые выявляются при вдумчивом авторском анализе [25].

Более серьезной является проблема фальсификации научного знания и существования псевдонауки, облегченной искусственным интеллектом. Так называемые «фабрики статей» (paper mills) уже начали использовать генеративный искусственный интеллект для штампования псевдонаучных публикаций. В данном контексте следует подчеркнуть, что генеративные языковые модели и программы для создания изображений позволяют относительно легко сочинять правдоподобные статьи с вымышленными данными, которые трудно сразу отличить от настоящих [26]. В 2023 г. мировое научное сообщество столкнулось с рекордным числом отозванных статей – свыше 10 000 работ были отозваны за один год, значительная часть – из-за потока фиктивных публикаций и мошенничества с рецензированием [27]. Крупные издательства вынуждены внедрять алгоритмы для выявления фальшивых рукописей. Так, издательство Wiley сообщало, что тестовый ИИ-детектор маркирует 10–13% поступающих рукописей как потенциально поддельные [28]. По оценкам А. Грау, уже от 1% до 17% научных статей 2023 г. могли быть написаны (полностью или частично) с помощью генеративных моделей [29]. Эти цифры свидетельствуют о начале массового проникновения машинного текста в академическую литературу.

С ростом доли научных текстов, созданных с использованием генеративных систем искусственного интеллекта, возникает фундаментальный вопрос о границах доверия к публикационной практике: станет ли научное сообщество по-прежнему воспринимать результаты исследований как эпистемически надежные, осознавая возможность их искусственного происхождения? В условиях, когда, по оценкам, каж-

дый пятый или даже десятый научный текст может быть синтезирован машинным алгоритмом, перед инфраструктурой доверия — включая институты экспертизы, систему рецензирования, академические библиотеки и базы данных — встает задача адаптации к новому когнитивному ландшафту, где наряду с традиционно полученным знанием циркулируют машинно-генерированные материалы.

Вместе с этим актуализируется угроза подрыва общественного доверия к научной деятельности в целом. Возникают ситуации, при которых ИИ-система, позиционируемая как эксперт, способна допустить ошибку — например, при диагностике заболевания или формулировании медицинских рекомендаций. Такие инциденты, ставшие достоянием широкой общественности, могут усиливать скептицизм в отношении научных экспертов, подрывая социальный авторитет науки и способствуя формированию климата недоверия в обществе. Аналитики вводят понятие «эпистемической безопасности общества» (epistemically secure society) – защиты общественной способности отличать знание от дезинформации [30; 31, с. 2].

В современных условиях искусственный интеллект становится ключевым вызовом в борьбе за эпистемическую безопасность научного знания. Диапазон угроз охватывает создание глубоко сфабрикованных материалов — от поддельных научных графиков и псевдоцитат, приписываемых признанным исследователям, до целенаправленных информационных кампаний, способных подрвать научный консенсус [32]. В этой связи формируется четвертая группа рисков, связанных с эрозией этических норм и снижением уровня доверия в научном сообществе. Отсутствие эффективных механизмов проверки, фильтрации и идентификации синтетического контента, генерируемого ИИ, существенно увеличивает вероятность инфляции публикационного шума. В условиях, когда научное информационное пространство насыщается автоматически созданными текстами и визуализациями, существенно усложняется задача различения достоверных результатов и их имитаций. Подобная динамика не только препятствует оперативной верификации новых знаний, но и ведет к размыванию границ между подлинной научной новизной и симулякрами интеллектуальной деятельности. Научное знание, по своей природе, интегрировано в коллективную инфраструктуру: его эволюция опирается на непрерывный диалог между исследователями, институтами экспертизы и обществом в целом. Подрыв доверия внутри этого сообщества способен поставить под угрозу не только эффективность академической коммуникации, но и саму возможность кумулятивного прогресса, лишая науку ее фундаментального свойства — способности к саморазвитию и критической рефлексии.

Таким образом, защита инфраструктуры доверия и этических оснований становится ключевым условием устойчивого развития научной среды в эпоху преобладания интеллектуальных алгоритмов.

## ЗАКЛЮЧЕНИЕ

В процессе проведенного исследования осуществлена комплексная систематизация эпистемологических угроз, обусловленных внедрением искусствен-

ного интеллекта в современную науку. Анализ эмпирических данных и теоретических источников позволил выявить основные направления трансформации научного знания под влиянием искусственного интеллекта; обосновать необходимость переосмысления традиционных принципов организации познавательной деятельности. Зафиксированы как количественные, так и качественные сдвиги в структуре академического дискурса, охватывающие изменения в логике производства, распространения и верификации научных результатов.

Выявлено, что эпистемологические риски связаны с утратой достоверности и воспроизводимости, усилением когнитивных искажений, а также снижением научного плюрализма и эрозией этических оснований исследовательской среды. Генеративные модели искусственного интеллекта оказываются способными не только ускорять и автоматизировать отдельные этапы научного поиска, но и способствовать институционализации новых форм знания, порой несущих угрозу для эпистемической устойчивости науки.

Обобщенные результаты подтверждают, что эффективное управление вызовами искусственного интеллекта требует разработки комплексных стратегий количественной и качественной оценки рисков, внедрения этических стандартов, а также институционализации практик, поддерживающих эпистемическое разнообразие и критическую рефлексию. Представленные выводы в совокупности, могут служить основой для дальнейших исследований в области философии и этики науки, а также для формирования политики ответственного использования искусственного интеллекта в академической среде.

Таким образом, представленная классификация эпистемологических угроз формирует концептуальные и практические ориентиры для повышения устойчивости, прозрачности и надежности научного знания в условиях динамического технологического развития.

## СПИСОК ЛИТЕРАТУРЫ

1. AI in science and research // *Fast Data Science*. — URL: <https://fastdatascience.com/ai-in-research/> (дата обращения: 05.05.2025).
2. Hao Q., Xu F., Li Y., Evans J. AI Expands Scientists' Impact but Contracts Science's Focus // *arXiv preprint arXiv:2412.07727*. — 2024. DOI: <https://doi.org/10.48550/arXiv.2412.07727>.
3. Peterson A.J. AI and the problem of knowledge collapse // *AI & Society*. — 2025. DOI: <https://doi.org/10.1007/s00146-024-02173-x>.
4. Филиппов О.А. Гносеологические риски цифровизации высшего образования: проблема верификации знания в условиях развития искусственного интеллекта // *Eurasian Journal of Social Sciences, Philosophy and Culture*. — 2025. — Vol. 5, № 3. — P. 113–119. DOI: <https://doi.org/10.5281/zenodo.15131787>.
5. Saeed W., Omlin C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities // *Knowledge-Based Systems*. — 2023. — Vol. 263, Art. 110273. DOI: <https://doi.org/10.1016/j.knosys.2023.110273>.
6. Duede E. Deep Learning Opacity in Scientific Discovery // *Philosophy of Science*. — 2023. — Vol. 90, № 5. — P. 1089–1099. DOI: <https://doi.org/10.1017/psa.2023.8>.
7. Balalle H., Pannilage S. Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity // *Social Sciences & Humanities Open*. — 2025. — Vol. 11. DOI: <https://doi.org/10.1016/j.ssaho.2024.100498>.
8. Papagiannidis E., Mikalef P., Conboy K. Responsible artificial intelligence governance: A review and research framework // *The Journal of Strategic Information Systems*. — 2025. — Vol. 34, №2. — Art. 101885. DOI: <https://doi.org/10.1016/j.jsis.2024.101885>.
9. OpenAI o3 and o4-mini System Card // OpenAI. — URL: <https://cdn.openai.com/pdf/2221c875-02dc-4789-800b-e7758f3722c1/o3-and-o4-mini-system-card.pdf> (дата обращения: 06.05.2025).
10. Иванова Л.А. Искусственный интеллект при написании научных статей — положительный или вредоносный фактор? // *Crede Experto: транспорт, общество, образование, язык*. — 2024. — № 4. — С. 6–17. DOI: [10.51955/2312-1327\\_2024\\_4\\_6](https://doi.org/10.51955/2312-1327_2024_4_6).
11. Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models // *United Nations Educational, Scientific and Cultural Organization (UNESCO)*. — URL: <https://unesdoc.unesco.org/ark:/48223/pf0000388971> (дата обращения: 08.05.2025).
12. OECD — *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*. — Paris: OECD Publishing, 2023. — 300 с. DOI: <https://doi.org/10.1787/a8d820bd-en>.
13. Messeri L., Crockett M.J. Artificial intelligence and illusions of understanding in scientific research // *Nature*. — 2024. — Vol. 627. — P. 49–58. DOI: <https://doi.org/10.1038/s41586-024-07146-0>.
14. Сычев А.А. В поисках надежности: трансформация доверия в эпоху цифровых технологий // *Мониторинг общественного мнения: экономические и социальные перемены*. — 2024. — № 5(183). — С. 37–59. DOI: [10.14515/monitoring.2024.5.2595](https://doi.org/10.14515/monitoring.2024.5.2595).
15. Holbrook C., Holman D., Clingo J. et al. Overtrust in AI Recommendations About Whether or Not to Kill: Evidence from Two Human-Robot Interaction Studies // *Scientific Reports*. — 2024. — Vol. 14. — P. 19751. DOI: <https://doi.org/10.1038/s41598-024-69771-z>.
16. Ullrich D., Butz A., Diefenbach S. The Development of Overtrust: An Empirical Simulation and Psychological Analysis in the Context of Human–Robot Interaction // *Frontiers in Robotics and AI*. — 2021. — Vol. 8, Art. 554578. DOI: [10.3389/frobt.2021.554578](https://doi.org/10.3389/frobt.2021.554578).
17. Лебедев С.А. Методологические проблемы истинности научного знания // *Журнал философских исследований*. — 2016. — Т. 2, № 6. — С. 2.
18. de Hond A.A.H., van Buchem M.M., Hernandez-Boussard T. Picture a data scientist: a call to action for increasing diversity, equity, and in-

- clusion in the age of AI // Journal of the American Medical Informatics Association (JAMIA). — 2022. — Vol. 29, № 12. — P. 2178–2181. DOI: <https://doi.org/10.1093/jamia/ocac156>.
19. Volle J., Schmitz A., Lietz H., Münch R. Group Formation in Science between Homogenization and Differentiation: Modeling the Development of U.S. and German Sociology // International Journal of Sociology. — 2024. — Vol. 54, № 4. — P. 221–241. DOI: <https://doi.org/10.1080/00207659.2024.2357908>.
  20. Кун Т. Структура научных революций / пер. с англ. — Москва: АСТ, 2003. — 320 с.
  21. Sendra-Balsells C., Campello V.M., Torrents-Barrena J. et al. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries // Scientific Reports. — 2023. — Vol. 13, № 1. — P. 2728. DOI: [10.1038/s41598-023-29490-3](https://doi.org/10.1038/s41598-023-29490-3).
  22. Celi L.A., Cellini J., Charpignon M.L., Dee E.C., Dernoncourt F., Eber R., Mitchell W.G., Moukheiber L., Schirmer J., Situ J., Paguio J., Park J., Wawira J.G., Yao S. Sources of bias in artificial intelligence that perpetuate healthcare disparities – A global review // PLOS Digital Health. — 2022. — Vol. 1, № 3. — Art. e0000022. DOI: <https://doi.org/10.1371/journal.pdig.0000022>.
  23. Kay J., Kasirzadeh A., Mohamed S. Epistemic Injustice in Generative AI // arXiv preprint arXiv:2408.11441. — 2024. — 23 p. DOI: <https://doi.org/10.48550/arXiv.2408.11441>.
  24. Тимохович А.Н., Самоходкина Е.Г., Самоходкин Е.В., Эльзон А.А. Проблематика цифровых технологий и искусственного интеллекта в научных работах // Цифровая социология. — 2023. — Т. 6, № 1. — С. 13-20.
  25. Cheng X., Zhang L. AI-generated literature reviews threaten scientific progress // Nature. — 2025. — Vol. 641, № 8064. — P. 852. DOI: [10.1038/d41586-025-01603-0](https://doi.org/10.1038/d41586-025-01603-0).
  26. Liverpool L. AI intensifies fight against ‘paper mills’ that churn out fake research // Nature. — 2023. — Vol. 618, № 7965. — P. 222–223. DOI: [10.1038/d41586-023-01780-w](https://doi.org/10.1038/d41586-023-01780-w).
  27. Van Noorden R. More than 10,000 research papers were retracted in 2023 — a new record // Nature. — 2023. — Vol. 624, № 7992. — P. 479–481. DOI: [10.1038/d41586-023-03974-8](https://doi.org/10.1038/d41586-023-03974-8).
  28. The paper mill problem: are AI tools the answer? // News for Medical Publication Professionals. — URL: <https://vk.cc/cMaRPS> (дата обращения: 12.05.2025).
  29. Gray A. ChatGPT "contamination": estimating the prevalence of LLMs in the scholarly literature // arXiv preprint arXiv:2403.16887. — 2024. DOI: [10.48550/arXiv.2403.16887](https://doi.org/10.48550/arXiv.2403.16887).
  30. Navigating the Complexities of Existential Risk: Insights from the 2023 Stanford Existential Risks Conference // Stanford Center for International Security and Cooperation. — URL: <https://vk.cc/cMaSFz> (дата обращения: 13.05.2025).
  31. Tackling threats to informed decisionmaking in democratic societies: Promoting epistemic security in a technologically-advanced world // The Alan Turing Institute: Defence and Security Programme. — URL: <https://vk.cc/cMaSS1> (дата обращения: 15.05.2025).
  32. Самоходкин Е.В., Эльзон А.А. Анализ взаимосвязи научного интереса и динамики публикаций по искусственному интеллекту в Российской Федерации (2020–2024 гг.) // Научно-техническая информация. Сер. 1. — 2025. — № 6. — С. 10–17. DOI: [10.36535/0548-0019-2025-06-2](https://doi.org/10.36535/0548-0019-2025-06-2).

*Материал поступил в редакцию 07.06.25.*

#### Сведения об авторах

**ТИМОХОВИЧ Александра Николаевна** – кандидат психологических наук, доцент Государственного университета управления, Москва  
e-mail: [3178720@list.ru](mailto:3178720@list.ru)

**САМОХОДКИН Евгений Вячеславович** – ведущий специалист Центра маркетинговых исследований и перспективного планирования Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН), Москва  
e-mail: [rodentforme@gmail.com](mailto:rodentforme@gmail.com)

**ЭЛЬЗОН Алиса Андреевна** – ведущий специалист Центра маркетинговых исследований и перспективного планирования ВИНИТИ РАН  
e-mail: [alisaelzon@gmail.com](mailto:alisaelzon@gmail.com)

# ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

---

УДК 004.6.08

Е.В. Бескаравайная, И.А. Митрошин, Н.А. Слащева

## Изучение пользовательских практик хранения данных

*Изложены результаты изучения мнений сотрудников Пушинского научного Центра относительно физических носителей, предназначенных для хранения информации, с целью: 1) определить информированность пользователей о технических средствах хранения информации; 2) выяснить использование устройств хранения данных на практике; 3) выявить достоинства и недостатки информационных носителей. Ответы, полученные на распространенные анкеты, позволили установить наиболее популярные и устаревшие устройства для оптимизации имеющегося оборудования и планирования приобретения современного технического и программного обеспечения.*

**Ключевые слова:** физические носители, устройства хранения данных, запоминающие устройства, носители данных, накопители, информационные носители, анкетирование пользователей

**DOI:** 10.36535/0548-0019-2025-11-2

### ВВЕДЕНИЕ

Сегодня спрос на создание высокопроизводительных сервисов хранения данных на основе новых быстрых устройств стал как никогда актуален. С одной стороны, это связано с потребностью сохранения больших объемов данных, с другой – с поиском возможностей организации длительного хранения, не зависящего от меняющихся технических и технологических составляющих. Публикации о носителях, предназначенных для хранения информации, как правило, делятся на две категории: технические статьи поставщиков (жестких дисков или устройств хранения данных) и описания исследований пользователей опыта. Поставщики дисков предоставляют информацию о том, как факторы окружающей среды или действия по использованию этих дисков могут влиять на срок их службы: описывают процессы и экспериментальную настройку для тестирования новых продуктов, пытаются сделать долгосрочные прогнозы надежности на основе ускоренных испытаний устройств. Нередко такая информация представляет собой рекламу услуг для сбыта продукции.

Цель нашего исследования – собрать информацию о физических носителях (оптические диски, магнитные ленты) и накопителях (HDD-диски, SSD – твердотельные накопители, USB-флешки и др.) потребительского класса, которыми на практике пользуются сотрудники институтов Пушинского научного Центра, выявить предпочтения пользователей, опреде-

лить положительные и отрицательные стороны носителей через призму их реального применения.

Актуальность обозначенной темы исследования связана с тем, что в библиотеках, архивах, информационных центрах накопились большие объемы материалов на разнообразных носителях (микрофиши, дискеты, магнитные ленты, жесткие диски, различные портативные накопители и пр.), и необходимо переосмыслить принципы их хранения с учетом меняющихся технических средств и программного обеспечения.

Объект исследования – устройства и методы хранения данных для организации сохранения и использования накопленных знаний вне зависимости от эволюционной смены программно-аппаратных методов хранения, обработки и доступа к данным.

Предмет нашего изучения – пользовательские практики применения различных типов устройств хранения данных сотрудниками Пушинского научного Центра.

Задачи исследования:

1) выявить предпочтения пользователей в выборе устройств для хранения данных, основываясь на опыте их эксплуатации;

2) определить спрос на персонализированные информационные носители для установки новых программных продуктов или сохранения устаревшего оборудования;

3) скорректировать на основе анализа потребностей пользователей имеющиеся технические и про-

граммные продукты в области хранения информации для организации действующего обслуживания и разработки стратегии длительного хранения.

Для сбора информации об использовании устройств хранения данных применялись методы анкетирования и интервьюирования, а также метод сравнения имеющихся технологий, возможностей технических и программных средств сервисов обслуживания пользователей на практике.

Практическую значимость настоящего исследования отражают следующие моменты:

- сбор и анализ применения устройств хранения данных позволяет выявлять наиболее значимые для пользователей носители информации;
- даже небольшое количество устаревших носителей предписывает обслуживающим организациям не отказываться от непопулярных операционных систем, средств просмотра и устройств хранения данных;
- появляющиеся накопители с новыми функциями делают актуальной задачу приобретения/коррекции современного технического и программного обеспечения.

## ОБЗОР ПУБЛИКАЦИЙ ПО ТЕМЕ

Публикации о тестировании физических носителей, в целом, описывают два направления исследований: 1) эксплуатация оборудования, т.е. устойчивость физических/технических параметров самих устройств; 2) надежность программного обслуживания, фокусирующаяся на поиске и отсуживании задержек, накоплении ошибок, прогнозировании отказов. Тестирование любых устройств сводится, как правило, к определению следующих параметров [1]:

- производительность (скорость чтения/записи, задержки, операций ввода/вывода в секунду (IOPS));
- надежность (устойчивость к перепадам напряжения, температурным колебаниям, вибрациям);
- долговечность (тестирование на износ);
- совместимость (работа с разными интерфейсами);
- безопасность (проверка шифрования, удаление данных без возможности восстановления и пр.).

Для нашего исследования мы выбрали наиболее, на наш взгляд, интересные или неоднозначные эксперименты по тестированию устройств хранения данных.

Еще в 1972 г. Национальное бюро стандартов США изучило воздействие различных опасностей на магнитные материалы, делая акцент на том, как сильные поля могут изменять или уничтожать данные; в опубликованном отчете предлагалось хранить носители на расстоянии не менее 8 см от трансформаторов, больших двигателей и другого оборудования, генерирующего поле [2]. Сотрудники миланского университета провели исследование микробной деградации – одного из наиболее серьезных и недооцененных источников повреждения компакт-дисков [3]. К факторам долговечности относится устойчивость к температурным колебаниям [4], однако опыт эксплуатации большого количества современных устройств сотрудниками Google Inc показал, что уровни температуры и активности гораздо меньше коррелируют с отказами дисков, чем сообщалось ранее [5]. Сотруд-

ники университета Карнеги-Меллона из Пенсильвании экспериментально доказали, что именно условия эксплуатации приводят к необходимости более ранней замены дисков, относительно гарантийного срока, чем факторы, связанные с компонентами [6].

Эволюция накопителей информации идет в сторону увеличения объема, скорости и долговечности хранения данных. На смену магнитным и оптическим носителям приходят флеш-накопители [7], сбои в работе которых уже прогнозируются на основании тестов искусственного интеллекта и машинного обучения [8].

Чтобы получить представление о проблемах и возможностях использования сетевых [9] и облачных хранилищ [10] с разнородными типами данных и сложными рабочими условиями, требуется мониторинг качества самих накопителей непосредственно в реальном времени [11] и заблаговременное прогнозирование отказов для предотвращения потери данных (SMART-анализ) [12].

При этом мы не встретили публикаций, анализирующие мнения обычных пользователей о носителях. В Интернете встречаются форумы с отзывами о том или ином накопителе для хранения цифровой информации, в том числе долговременного, однако, как правило, это оценки профессионалов, работающих в сфере информационных технологий: программистов, системных администраторов и разработчиков программного обеспечения. Мы же работаем с пользователями, различающимися как по возрасту, так и по уровню владения персональным компьютером, для которых выбор устройства хранения данных зависит от множества факторов: стоимости, удобства использования, размера, веса, просто привычки или способа получения (в подарок, с материалами конференции и т.д.). Именно поэтому, анализ информации о наличии запоминающих устройств и их оценке нашими пользователями стал для нас основой организации обслуживания.

## МОНИТОРИНГ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИОННЫХ НОСИТЕЛЕЙ

Исследования опыта применения носителей может не иметь такой глубины понимания внутренних механизмов надежности устройств, как это возможно в отчетах производителей, но это необходимо нам для понимания использования устройств в реальных условиях непосредственно пользователями.

Нами была разработана анкета с перечнем запоминающих устройств для хранения информации, в которой предлагалось ответить на вопросы, включающие сведения о сроках использования тех или иных устройств хранения данных; типах информации, которая на них сохраняется; причинах отказа от использования этих устройств и др. Анкеты были направлены респондентам по электронной почте, предложены в распечатанном виде, представлены на сайте библиотеки. Всего было распространено 450 анкет, из которых с ответами вернулись 417 (92,6%).

Возрастной состав респондентов распределился следующим образом: до 30 лет – 18%, 30-60 лет – 71%, более 60 лет – 11%. Среди них: профессиональ-

ное отношение к IT-технологиям имели 24 человека (системные администраторы, программисты, специалисты на стыке информационных технологий и естественно-научных областей (биоинформатики, аналитики и пр.); аспиранты, магистранты и студенты – 58 человек; научные сотрудники (младшие, старшие, ведущие) – 295 человек; главные научные сотрудники, руководители подразделений, администрация институтов – 26 человек; 14 человек составили группу, не имеющую отношение к научной работе (инженеры, сотрудники библиотеки, обслуживающий персонал). По данным наших предыдущих опросов такое распределение, в целом, соответствует возрастному и должностному составу Пущинского научного Центра.

На начальном этапе исследования респондентам были представлены описания наиболее распространенных запоминающих устройств, чтобы создать у них четкое понимание имеющихся накопителей и носителей для хранения информации (табл. 1).

Затем респондентам было предложено отметить все запоминающие устройства, с которыми они имеют опыт работы и выбрать те, которые они используют для исследовательских и личных целей на современном этапе (рис. 1).

Наиболее актуальными, как по опыту использования, так и по современному применению, остаются HDD (жесткие диски), USB-флешки, переносные жесткие диски и облачные хранилища. При этом, на сегодня богатый опыт применения имеют ленточные накопители (82%) и оптические диски всех типов (более 60%). Как правило, в рабочих целях, например, на магнитных лентах хранятся записи голосов птиц в фонотеке Б.Н. Вепринцева, а на CD и DVD-носителях в лабораториях архивируются исследовательские материалы, данные экспериментов, библиографии сотрудников, оригинальное программное обеспечение/установочные данные для операционных систем.

Таблица 1

**Наиболее распространенные виды информационных носителей**

<b>Носитель</b>	<b>Описание/формат</b>
HDD - жесткие диски (Hard Disk Drive)	Используют магнитные диски (пластины) для хранения данных; объем от нескольких сотен гигабайт до нескольких терабайт; размер – 3,5 дюйма для настольных компьютеров и 2,5 дюйма для ноутбуков; подключаются через интерфейсы SATA (Serial ATA) и более старые интерфейсы, такие как IDE (PATA)
USB-флешки	Энергонезависимые хранилища, объем от нескольких сотен гигабайт до нескольких терабайт; скорость передачи зависит от версии USB (например, USB 2.0, USB 3.0, USB 3.1, USB 3.2), совместимы с большинством устройств, имеющих USB-порт
SSD (твердотельные накопители)	Используют микросхемы флеш-памяти, высокую скорость чтения и записи данных; устойчивы к механическим повреждениям, ударам и вибрациям; размеры варьируют от 2,5-дюймов для ноутбуков и настольных компьютеров, и компактные M.2 и mSATA для ультрабуков и планшетов; объем от нескольких сотен гигабайт до нескольких терабайт; используют различные интерфейсы подключения (SATA, PCIe и NVMe), имеют ограниченное количество циклов записи. Портативные накопители (внешние SSD) подключаются через USB или Thunderbolt
Гибридные системы SSHD (Solid State Hybrid Drive)/	Объединяют в одном устройстве магнитные пластины для хранения больших объемов данных и флеш-память для ускорения доступа к часто используемым данным; высокая скорость загрузки
Ленточные носители (Linear Tape-Open)	Объемы данных от нескольких терабайт до десятков терабайт на одну кассету; используются для резервного копирования и архивирования; поддерживают шифрование данных
CD-ROM (Compact Disc Read-Only Memory)	Оптический носитель информации, изготавливается из поликарбонатного пластика с металлическим покрытием для отражения лазерного луча; данные объемом не более 700 МБ, считываются с помощью лазера, информация не может быть изменена или удалена
CD-R (Compact Disc Recordable)	Оптический носитель информации, данные записываются на CD-R с помощью лазера; информация не может быть изменена или удалена; совместимы с большинством CD-плееров и приводов; стандартная ёмкость 700 МБ; может потребоваться специальное программное обеспечение
CD-RW (Compact Disc ReWritable)	Оптический носитель информации, перезаписываемый диск, скорость записи может варьироваться в зависимости от привода; стандартная ёмкость 700 МБ; старые модели приводов могут испытывать трудности с чтением или записью CD-RW; может потребоваться специальное программное обеспечение
DVD-ROM (Digital Versatile Disc Read-Only Memory)	Оптический носитель информации, записанные данные не могут быть изменены; может хранить от 4,7 ГБ до 8,5 ГБ (двухслойный); совместимы с большинством DVD-плееров и компьютерных DVD-приводов, не совместимы с CD-плеерами и приводами; используется для распространения видео и программного обеспечения

Носитель	Описание/формат
DVD-R/DVD+R:	Оптический носитель информации, поддерживает однократную запись данных, скорость записи может варьироваться в зависимости от привода и диска, имеет высокую степень совместимости с большинством DVD-плееров и приводов, объем 4,7 Гб для однослойного диска и 8,5 Гб для двухслойного диска; DVD+R имеет лучшее управление записью и возможность записи в несколько сессий
DVD-RW/DVD+RW	Оптический носитель информации, перезаписываемые DVD-диски; скорость записи варьируется в зависимости от привода и диска; старые устройства могут испытывать трудности с их чтением; ёмкость 4.7 Гб; в DVD+RW улучшенное управление записью и возможность записи в несколько сессий
HD DVD	Оптический носитель, разработанный для хранения видео высокой четкости (HD); использует синий лазер для чтения и записи данных; для чтения использует HD DVD приводы; емкость однослойных - 15 Гб данных, а двухслойный — до 30 Гб; использует систему защиты контента AACS (Advanced Access Content System); снят с производства
Оптические диски (BD-R) (Blu-ray Disc Recordable)	Оптический носитель, предназначенный для однократной записи данных; высокой плотности хранения по сравнению с DVD и CD; однослойный BD-R диск может хранить до 25 Гб данных, четырехслойный 128 Гб; скорость записи варьируется в зависимости от привода и диска; используются для хранения данных и видео высокой четкости; совместимы с большинством Blu-ray плееров и приводов
Оптические диски (M-DISC)	Оптический носитель с неорганическим записывающим слоем, более устойчивый к воздействию окружающей среды и времени; совместим с большинством современных DVD и Blu-ray приводов и плееров; доступен в различных форматах, включая DVD и Blu-ray; емкость 25 Гб для однослойного и до 50 Гб для двухслойного диска
Сетевые хранилища (NAS — Network Attached Storage)	Обеспечивают доступ к данным нескольким пользователям и устройствам одновременно; поддерживают: возможность добавления дополнительных дисков или расширения ёмкости, функции потоковой передачи мультимедийного контента, различные сетевые протоколы, такие как SMB/CIFS, NFS, AFP и др.
Облачные хранилища (т.к. (Google Drive, Dropbox, iCloud))	Позволяют пользователям: хранить, управлять и получать доступ к своим данным на удаленных серверах, работать с файлами и данными на различных устройствах, таких как компьютеры, смартфоны и планшеты, легко масштабируются; позволяют нескольким пользователям одновременно работать с одними и теми же файлами и данными; поддерживают автоматическое обновление программного обеспечения и функций; интегрируются с другими облачными сервисами и приложениями: офисные пакеты, инструменты для совместной работы и аналитики; поддерживают резервное копирование данных и их восстановления в случае потери или повреждения локальных копий

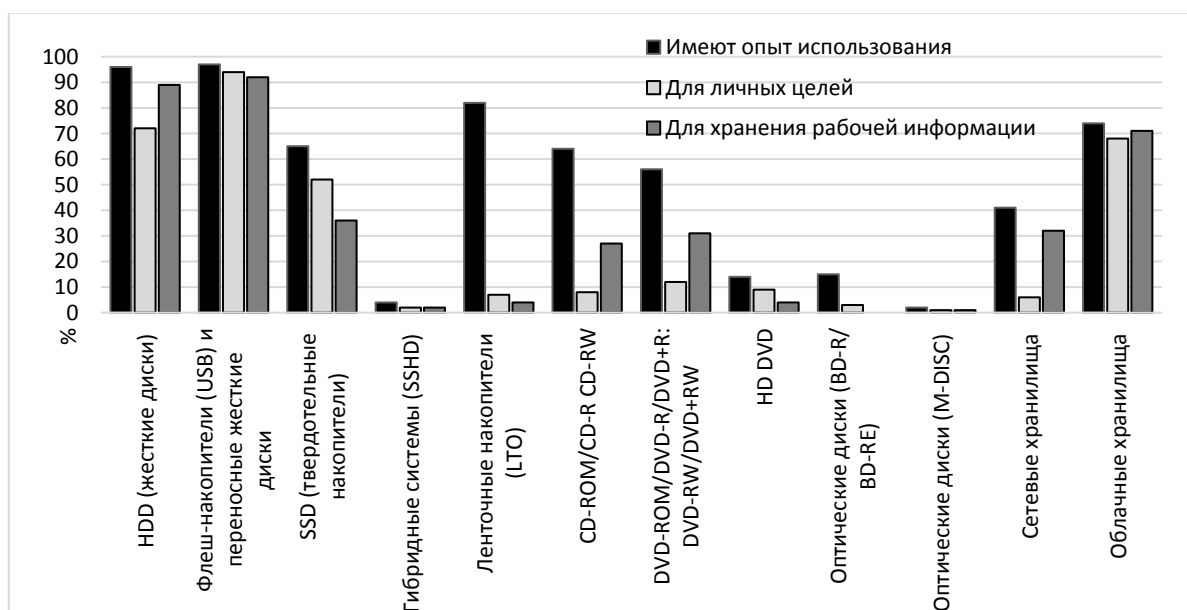


Рис. 1. Распределение информационных носителей по их использованию респондентами

Следует отметить заметную корреляцию между типом применяемого устройства и возрастом респондента. Например, в ответах возрастной категории до 30 лет практически не фигурируют магнитные ленты или оптические диски; эти респонденты делают акцент на SSD и возможностях облачных технологий. Напротив, люди старшего поколения отметили, что продолжают хранить в домашних архивах аудиозаписи на магнитных лентах и фото/видеозаписи на оптических дисках, а для рабочих данных используют переносные устройства (жесткие диски и флешки). Наиболее разнообразными были ответы возрастной категории от 30 до 60 лет. Эти пользователи не только видели развитие технологии хранения, но и испытали на себе проблемы, возникавшие из-за устаревания носителей, способов записи, прекращения поддержки со стороны производителей программного обеспечения и операционных систем. Как следствие – их осведомленность в вопросах создания резервных копий.

Организуя длительное хранение, пользователи предпочитают разделять запоминающие устройства по типам сохраняемых файлов. Как видно на рис. 2, наиболее универсальными решениями в этом плане являются жесткие диски, твердотельные накопители, сетевые и облачные хранилища, а магнитные ленты и оптические диски, как правило, служат архивом для медиа файлов.

Концепция «длительного сохранения» на данный момент разработана довольно слабо, начиная от механизмов отбора данных, до отсутствия достаточного опыта жизнеспособности отдельных носителей. Компромиссом остается регулярное резервное копирование. Более половины участников опроса подтвердили, что создают копии своих данных регулярно или время от времени, выбирая для их хранения внутренние и переносные жесткие диски и SSD, сетевые хранилища (рис. 3), а часть респондентов указали, что процесс создания копии (backup) у них настроен автоматически по заданному расписанию.

Многолетний опыт поднимает важный вопрос о причинах отказа пользователей от какого-либо типа информационных носителей. В анкете предлагалось 6 наиболее распространенных доводов. Основные из них, подтолкнувшие респондентов отказаться от использования:

- физические повреждения (для HDD, USB-флешек и оптических дисков);
- проблемы с оборудованием для чтения/записи (для ленточных накопителей и оптических дисков);
- безопасность хранимых данных;
- развитие/смена технологий (для USB-флешек, оптических дисков, магнитных лент);
- низкая скорость работы (для HDD и магнитных лент);
- стоимость (для SSD и сетевых хранилищ).

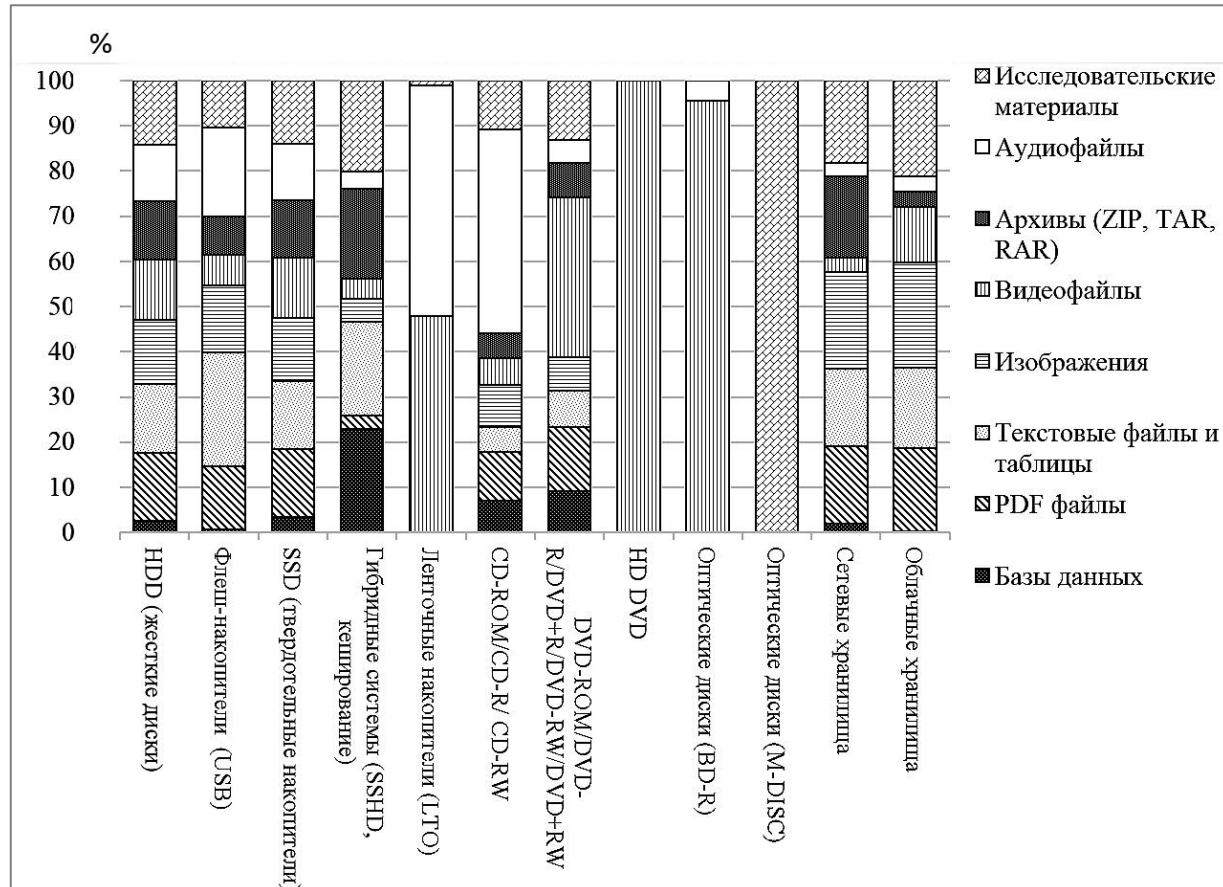


Рис. 2. Выбор запоминающих устройств для длительного хранения в зависимости от типа сохраняемых файлов

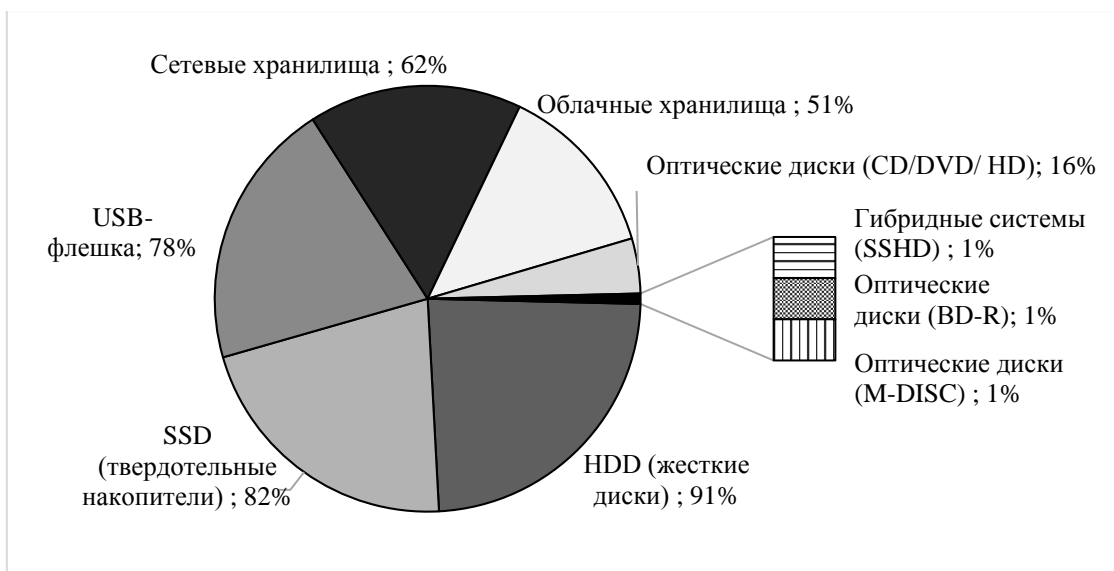


Рис. 3. Использование различных типов носителей в организации резервного копирования (процент подсчитан от количества положительных ответов о резервном копировании)

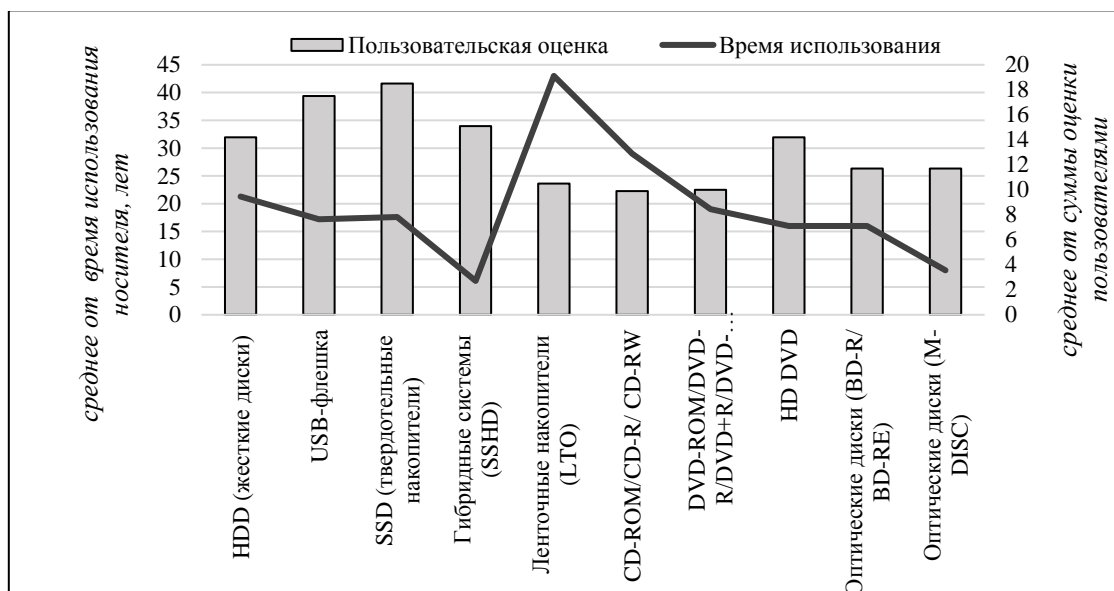


Рис. 4. Оценка устройств хранения данных пользователями

Таблица 2

#### Планирование респондентами использования устройств хранения информации

Вид носителя	Уже использую и планирую в дальнейшем, %	Использую, не планирую в дальнейшем, %	Не использую, но планирую в дальнейшем, %	Не использую и не планирую, %
HDD (жесткие диски)	89	7		4
USB-флешка	68	22	1	9
SSD (твердотельные накопители)	60	5	34	1
Гибридные системы (SSHD)	3	12	1	84
Ленточные накопители (LTO)		14		86
Оптические диски (CD/DVD/HD/BD-R)	2	32		66
Оптические диски (M-DISC)	1			99
Сетевые хранилища	24	0	7	69
Облачные хранилища	60	8	11	21

8% опрошенных пользователей, имеющих практику работы с облачными технологиями (74%), отметили проблемы, возникающие при их использовании, что послужило отказом от применения облачных хранилищ:

- ограничение бесплатного места – 70%;
- потеря файлов – 7%;
- ограниченная скорость загрузки/выгрузки и утечка данных – по 4%;
- невозможность загрузить/синхронизировать данные, удаление файлов, автоматическая блокировка файлов из-за цензуры – по 2%;
- взлом и блокировка аккаунта – 1%.

Кроме того, в качестве недостатков, пользователи отметили трудности синхронизации данных при медленном интернет-соединении, а некоторые упомянули скрытую рекламу и навязчивый маркетинг.

Участники анкетирования достаточно четко определились с дальнейшим выбором для себя систем хранения (табл. 2).

В ответ на вопрос об устройствах для хранения данных, большинство респондентов оставили в приоритете жесткие диски, SSD диски (внешние и портативные) и облачные хранилища, при этом отказаться от оптических дисков планируют 32% пользователей, а от USB-флешек – 22%.

В заключительной части анкеты было предложено, основываясь на личном опыте, оценить по пятибалльной шкале каждое запоминающее устройство по критериям:

- скорость чтения/записи,
- задержки
- среднее время до замены,
- устойчивость к экстремальным условиям» (температура, вибрация, влажность, скачки напряжения),
- цена,
- объем,
- удобство использования.

Из полученных ответов мы составили представление о субъективных оценках информационных носителей обычными пользователями. Первое место было отдано твердотельным накопителям SSD – они получили наивысший средний балл как наиболее надежные, быстрые и энергоэффективные; HDD – были положительно отмечены с точки зрения хранения больших объемов, USB-флешки – как самые мобильные и дешевые, а гибридные системы – как сочетающие скорость и объем (рис. 4).

По функционалу устройств мнения пользователей, в целом, были сходными:

- если необходимо сохранять большой объем данных или создавать резервные копии, то лучше подходит HDD;
- если необходима скорость, то следует использовать SSD (но дорого за большой объём);
- если требуется быстро перенести данные с устройства на устройство, то выбор стоит остановить на USB-флешки и внешних SSD (но не для хранения критически важного).
- для сохранения данных в качестве резервных копий лучше всего подходят облачные технологии, локальный HDD, а на работе – сетевое хранилище;

- для создания долгосрочного архива, который не требует постоянного доступа и поиска, можно применять оптические диски типа Blu-ray или M-Disc;

- сетевые хранилища наиболее функциональны для контроля за данными и быстрого доступа к большим данным, даже при медленном Интернете; удобны для подключения к общему оборудованию (свитчи, маршрутизаторы, принтеры и др); имеют возможность автоматической интеграции данных из сетевых хранилищ в облако для резервного копирования.

## ВЫВОДЫ

Анализ применения носителей для хранения информации позволяет сделать следующие выводы.

1. Несмотря на то, что жёсткие диски (HDD), по сравнению с SSD, «...шумят, медленнее загружаются и боятся ударов», их дешевизна и большая ёмкость оставляют их важным устройством для хранения больших объемов информации на практике. Именно на этих накопителях пользователи хранят все виды документов, включая их резервные копии. Из недостатков HDD пользователи отметили чувствительность к перегреву и резким скачкам напряжения, а IT-специалисты подчеркнули необходимость периодической дефрагментации для предотвращения замедления работы.

2. При оценке USB-флешки большинство пользователей указывают на ненадежность этого носителя, предпочитают использовать его для переноса информации, но никак не для длительного хранения; тут играет свою роль и небольшая емкость, и медленная скорость переноса файлов. Тем не менее, на сегодняшний день – это самый распространенный накопитель у респондентов, особенно старшего поколения, на нем они хранят как личные документы, так и материалы исследований.

3. Что касается оптических дисков всех типов, то в ежедневной работе они используются довольно редко, это обусловлено необходимостью наличия дисководов, быстрым повреждением записывающего слоя при частой эксплуатации. В то же время большинство пользователей сохраняет на них информацию (фото и видео файлы, документы, диссертации, личную библиографию и др.) в качестве архивов. Более того, как выяснилось из опроса, часть респондентов, оцифровывала информацию, хранящуюся на магнитных лентах, переписав ее на оптические диски – отсюда такая высокая доля их участия в ответах. Магнитные же ленты, несмотря на то, что это один из самых ёмких и энергонезависимых носителей, и, на данный момент, они широко применяются в мировой практике для холодного архивирования больших данных, пользователями не задействуются. На наш взгляд, это связано не только с необходимостью применять специальное оборудование, но и невозможностью осуществлять быстрый поиск на них.

4. Как видно из ответов на вопросы анкеты, гибридные накопители SSHD оказались непопулярны среди простых пользователей: 84% опрошенных не используют и не планируют использовать их в дальнейшем, упоминая, что для работы быстрее SSD, а для хранения больших объемов надежнее HDD. Они

остаются пока в недорогих моделях ноутбуков, которые встречаются у респондентов, но на смену им приходят модели на SSD.

5. Мнения пользователей о твердотельных накопителях (SSD) самые положительные – это касается как стационарных, так и переносных SSD: скорость, надежность, большой объем и компактность, при этом все отмечают слишком высокую цену, а IT-специалисты еще и трудность восстановления данных при поломке. По мнению наших пользователей, это самый перспективный накопитель как для работы, так и для хранения информации. Среди опрошенных 64% уже используют их в работе, а 32 % собираются перейти на них в будущем.

Следует отметить, что наряду с производительностью и доступностью самих запоминающих устройств, на их распространенность влияют возраст и привычки пользователей. Например, представители старшего поколения не собираются отказываться от USB-флешки в пользу более быстрых внешних SSD, а молодые сотрудники готовы поработать с новыми технологиями, если они появятся.

В зависимости от своих потребностей пользователи по-разному оценивают различные типы информационных носителей, и в этом немалую роль играет их стоимость. Принимавшие участие в наших исследованиях системные администраторы и IT-специалисты, высказали интересное замечание: как правило, в личных целях приобретаются SSD более дорогого и надежного сегмента, и время их безотказной работы составляет в среднем 6,5 лет, тогда как в учреждениях из-за экономии средств, предлагаются более дешевые варианты, средний срок службы которых варьируется от 1,5 до 3 лет.

Одной из задач нашего анкетирования было выявление у пользователей устаревших систем хранения. С появлением CD, DVD, флешек и облачных хранилищ, дискеты быстро потеряли актуальность, тем не менее их портативность, простота использования, дешевизна и возможность многократно стирать и записывать данные, оставляла их редким, но применяющимся на практике накопителем. Для пользователей дискет на одном из компьютеров был установлен внешний USB-дисковод для гибких дисков (FDD — Floppy Disk Drive) и специализированное программное обеспечение для чтения старых форматов файлов. На данный момент респонденты не указали в анкетах факт использования дискет, и мы можем отказаться от дисководов для гибких дисков.

Несколько слов о сетевых хранилищах. Опыт их применения имеют 41% респондентов, при этом большинство опрошенных (32%), используют их для хранения исследовательских материалов и обмена данными в локальной сети на работе. Личное использование включает хранение фото/видео из семейных архивов, резервное копирование документов с целью доступа к ним с других устройств (телефонов, ТВ, ноутбуков), доступ к общему оборудованию: например, IP-камерам, системе «умный дом» и др.).

Что касается облачного хранения, то наиболее активно задействованы Google Drive, Yandex Disk, Облако Mail как оптимальные по удобству и объему. В их пользу респонденты отказались от практики

OneDrive и Dropbox из-за небольшого объема бесплатного хранилища. Самыми безопасными указаны Proton Drive, MEGA, pCloud, однако хранение в них дорого и требует дополнительных навыков. Часть пользователей применяют стратегию трехкратного резервного копирования, при которой одинаковые версии хранятся на рабочем компьютере, внешнем носителе и в облаке. В процессе сбора информации мы выявили интересный момент: активно работая с системами Google, Яндекс, Mail, не все опрошенные осознают, что они автоматически используют и облачные хранилища этих сервисов.

Наши результаты показывают, что на данный момент пользователи не представили технических продуктов и технологий, для которых у нас нет соответствующего оборудования и программного обеспечения в области хранения/визуализации информации.

Одним из наиболее вероятных сценариев в ближайшее время станет массовое внедрение новых технологий, например, распространение USB4, поэтому, планируя сегодня закупку оборудования, мы должны учитывать его совместимость с Thunderbolt и DisplayPort, наличие портов USB-C (Type-C), кабелей USB4 и т.д.

Таким образом, одним из важных направлений развития систем хранения документов в библиотеке становится перенос данных с устаревающих устройств, операционных систем и форматов на современные. В качестве примера можно привести материалы, хранимые у нас в виде микрофишей – устаревшей, но живой технологии, работа с которой требует специальных аналоговых читальных аппаратов. В течение последних 22-х лет не было ни одного обращения за ними в библиотеку, и микрофиши не были упомянуты ни в одной из анкет. Исходя из этого, мы сделали вывод, что данный носитель неудобен для использования и не востребован пользователями, поэтому материал на микрофишах мы планируем оцифровать, а сам информационный носитель отправить в архив, соответствующий требованиям длительного хранения.

## СПИСОК ЛИТЕРАТУРЫ

1. Вальков В.А., Столяров Е.П., Корчагин А.А., Ермишин М.В., Якупов Д.О. Сравнение методов оптимизации скорости чтения/записи накопителей // Программные системы и вычислительные методы. – 2024. – № 2. – С. 73-85. – URL: <https://cyberleninka.ru/article/n/sravnenie-metodov-optimizatsii-skorosti-chteniya-zapisi-nakopiteley> (дата обращения: 26.07.2025).
2. Katzive B. Matching Magnetic Media with Modern Machines // Digital Design. – 1977. – Vol. 7, № 5. – P. 20-34.
3. Cappitelli F., Sorlini C. From Papyrus to Compact Disc: The Microbial Deterioration of Documentary Heritage // Critical Reviews in Microbiology. – 2005. – Vol. 31(1). – P. 1–10. DOI: <https://doi.org/10.1080/10408410490884766>
4. Sankar S., Shaw M., Vaid K., Gurumurthi S. Datacenter scale evaluation of the impact of temperature on hard disk drive failures // ACM Transactions

- on Storage (TOS). – 2013. – Vol. 9, № 2. – P. 1-24. DOI: <https://doi.org/10.1145/2491472.2491475>
5. Pinheiro E., Weber W.D., Barroso L.A. Failure Trends in a Large Disk Drive Population // Fast. – 2007. – Vol. 7, №. 1. – P. 17-23.
  6. Schroeder B., Gibson G.A. Understanding disk failure rates: What does an MTTF of 1,000,000 hours mean to you? // ACM Transactions on Storage (TOS). – 2007. – Vol. 3, №. 3. – P. 8. DOI: <https://doi.org/10.1145/1288783.1288785/>
  7. Белова Е.Ю. Обзор систем хранения данных на флеш-накопителях // Известия СПбГЭТУ ЛЭТИ. – 2017. – № 9. – С. 27-33.
  8. Song Y., Liang Y., Liu J., Shi L. Prophet: SSD Failure Analysis and Prediction Guided by Flash Reliability Characteristics in Data Centers // IEEE Transactions on Computers. – Aug. 2025. – Vol. 74, Iss. 8. – P. 2529-2541. DOI: 10.1109/TC.2025.3566871.
  9. Колтышев В.В., Уймин А.Г. Обзор и сравнительное тестирование сетевых хранилищ // Фундаментальные и прикладные исследования: проблемы и результаты. – 2013. – № 4. – С. 164-169. – URL: <https://cyberleninka.ru/article/n/obzor-i-sravnitelnoe-testirovanie-setevyh-hranilisch> (дата обращения: 05.08.2025).
  10. Yang F., Yin W., Wang L., Li T., Zhao, P., Liu B. et al. Diffusion-based time series data imputation for cloud failure prediction at Microsoft 365 // Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. FSE'23 Industry, San Francisco – 2023. – P. 2050-2055.
  11. Chen Y., Yang X., Lin Q., Zhang H., Gao F., Xu Z., et al. Outage prediction and diagnosis for cloud service systems // The World Wide Web conference. – 2019. – P. 2659-2665. DOI: <https://doi.org/10.1145/3308558.3313501>
  12. Tomer V., Sharma V., Gupta S., Singh, D.P. Hard disk drive failure prediction using SMART attribute // Materials Today: Proceedings. – 2021. – Vol. 46. – P. 11258-11262.

*Материал поступил в редакцию 07.08.25.*

#### **Сведения об авторах**

**БЕСКАРАВАЙНАЯ Елена Вячеславовна** – старший научный сотрудник, Библиотека по Естественным Наукам Российской академии наук (БЕН РАН), Москва  
e-mail: [elenabesk@gmail.com](mailto:elenabesk@gmail.com)

**МИТРОШИН Иван Андреевич** – старший научный сотрудник БЕН РАН  
e-mail: [imitros@gmail.com](mailto:imitros@gmail.com)

**СЛАЩЕВА Наталья Анатольевна** – кандидат педагогических наук, старший научный сотрудник БЕН РАН  
e-mail: [slashcheva@rambler.ru](mailto:slashcheva@rambler.ru)

## Практика формирования метаданных о научной деятельности: анализ сайтов и открытых институциональных информационных систем\*

*Анализ метаданных 52-х открытых российских платформ (сайты, институциональные информационные системы текущих исследований) проведен на четырех основных объектах: профиль автора, публикации, научные проекты и награды. Для каждого объекта был выявлен определенный набор характеризующих его основных и факультативных элементов. Отмечается, что наиболее полный набор метаданных результатов научно-исследовательской деятельности формируется в информационных системах текущих исследований. Значительное внимание уделяется формированию метаданных, раскрывающих квалификацию сотрудника, его профессиональную, учебную и научную деятельность.*

**Ключевые слова:** открытая наука, метаданные, информационные системы текущих исследований, CRIS-системы

**DOI:** 10.36535/0548-0019-2025-11-3

### ВВЕДЕНИЕ

В условиях высокой конкуренции повышаются требования, предъявляемые к оценке эффективности результатов, получаемых в ходе научных исследований. Для сбора информации о научных разработках, проводимых в учреждениях и их структурных подразделениях (направления, проекты, финансирующие организации, публикационная активность и т. п.), а также о сотрудниках (персональные данные ученых) разрабатываются автоматизированные информационные системы Current Research Information Systems (CRIS-системы).

Изучение возможностей CRIS-систем позволяет выделить перспективы их применения: для обогащения метаданных и создания эффективной коммуникации между учеными и исследовательскими коллективами [1]. Многие авторы акцентируют внимание на качестве метаданных [2, 3], отмечая, с одной стороны, их важность для применения с целью оценки эффективности научно-исследовательской деятельности, от результатов которой зависит финансирование университетов, и, с другой стороны, их влияние на продвижение имиджа ученого и научной организации, обеспечивая видимость результатов деятельности в сети. Подчеркивается, что данные должны быть доступны, повторяемы и актуальны [4].

Цель настоящего исследования – изучение качества метаданных, формируемых на площадках от-

крытых российских институциональных информационных систем текущих исследований (ИИСТИ) и сайтах научных учреждений.

Для изучения качества метаданных (полнота и корректность их формирования) о результатах научно-исследовательской деятельности учреждения были рассмотрены открытые российские платформы (сайты, ИИСТИ). Всего проанализировано 52 учреждения, из них 13 университетов и 39 научно-исследовательских институтов, которые осуществляют сбор информации о научных результатах на открытых площадках в сети.

### CRIS-СИСТЕМЫ И ИХ РАЗНОВИДНОСТИ

Размещение данных в сети способствует выполнению определенных задач: информирование о результатах научной деятельности, формирование имиджа ученого и/или научного учреждения, налаживание научных коммуникаций, доступ к исследовательским данным в любое время и в любом месте. Среди существующих видов CRIS-систем (континентальные, национальные, региональные, институциональные) самыми распространенными являются институциональные системы. Первые три вида имеют единичные мировые примеры. Так, континентальная информационная система EuroCRIS была создана в 2000 г. и является результатом деятельности Международной профессиональной ассоциации в направлении объединения разработчиков научных информационных систем и их исследователей, на ее платформе собрана информация о CRIS-системах разных европейских стран, в том числе и России. Членами euroCRIS по состоянию на 2025 г. являются 7 CRIS-систем сле-

\* Статья подготовлена по плану НИР ГПНТБ СО РАН, проект «Разработка модели функционирования научной библиотеки в информационной экосистеме открытой науки», № 122041100150-3.

дующих российских организаций: Российской академии наук, Российского университета дружбы народов, Санкт-Петербургского государственного университета, Уральского федерального университета, Томского политехнического университета, Новосибирского государственного университета. Кроме того, на платформе EuroCRIS зарегистрированы информационные системы Индии, Норвегии, Великобритании, Италии, Испании, США, Германии и других стран (табл. 1). Регистрационные данные свидетельствуют о постепенном увеличении общего числа CRIS-систем, особенно это характерно для Индии. За несколько лет число зарегистрированных на платформе EuroCRIS систем увеличилось до 631(+113).

Таким образом, внедрение и разработка CRIS-систем приобрела сегодня мировой масштаб, накоплен существенный опыт их применения для учета результатов научно-исследовательской деятельности. Помимо EuroCRIS разработаны информационные системы национального уровня в Норвегии, Бельгии, Италии, Германии, Исландии, Дании, Швеции и других европейских стран [5-10].

В России информация об интеллектуальной деятельности исследователей формируется на сайтах учреждений и создаваемых ими специализированных информационных платформах – CRIS-системах. Одним из первых примеров стала научная информационная система «Соционет» (<https://socionet.ru/>), которая была создана с использованием результатов и рекомендаций международных инициатив в рамках программы «Открытый доступ к результатам исследований» [11]. В ее разработке принимали участие специалисты из разных институтов Российской академии наук. Однако в апреле 2022 г. проект был закрыт на основании решения суда по причине нарушения авторских прав.

Другим крупным российским проектом национального масштаба [12] стала Интеллектуальная система тематического исследования наукометрических данных «ИСТИНА» Московского государственного университета им. М.В. Ломоносова (<http://istina.msu.ru/>). На текущий момент система активно развивается, на ее платформе размещены результаты научно-исследовательской деятельности 34-х организаций, среди них 7 университетов (МГУ им. М.В. Ломоносова, Российский химико-технологический университет им. Д.И. Менделеева, Московский педагогический государственный университет, Псковский государственный университет, Санкт-Петербургский государственный химико-фармацевтический университет, Волжский государственный университет и Российский государственный гуманитарный университет) и 27 научно-исследовательских институтов. На платформе размещены метаданные о разных типах материалов (статьи, доклады на конференциях, тезисы докладов, книги, диссертации и т. д.). Практически половину от общего количества публикаций составляют статьи – более 1 млн записей – 49 % (рис. 1).

Многие научно-исследовательские учреждения создают собственные проекты CRIS на базе программного обеспечения (ПО) с открытым исходным кодом (например, DSpace-CRIS). Информационная платформа DSpace, в свою очередь, позволяет встраивать ограниченный набор элементов MARC-записи [13]. В качестве основы для регистрации и учета результатов научно-исследовательской деятельности учреждения в информационных системах используются базы данных трудов сотрудников. Несколько проектов таких информационных систем реализовано в учреждениях Российской академии наук [14-18].

Таблица 1

**География специализированных информационных платформ, зарегистрированных в EuroCRIS (топ-10 стран)**

Страна	Количество зарегистрированных CRIS в 2023 г.	Страна	Количество зарегистрированных CRIS в 2025 г.	Изменение количества зарегистрированных CRIS за два года
Индия	518	Индия	631	+113
Норвегия	136	Норвегия	136	0
Великобритания	102	Великобритания	104	+2
Италия	80	Италия	82	+2
Германия	59	Испания	76	+10
США	50	Германия	69	+18
Испания	46	США	68	+30
Польша	43	Польша	57	+14
Турция	37	Турция	38	+1
Нидерланды	28	Нидерланды	34	+6

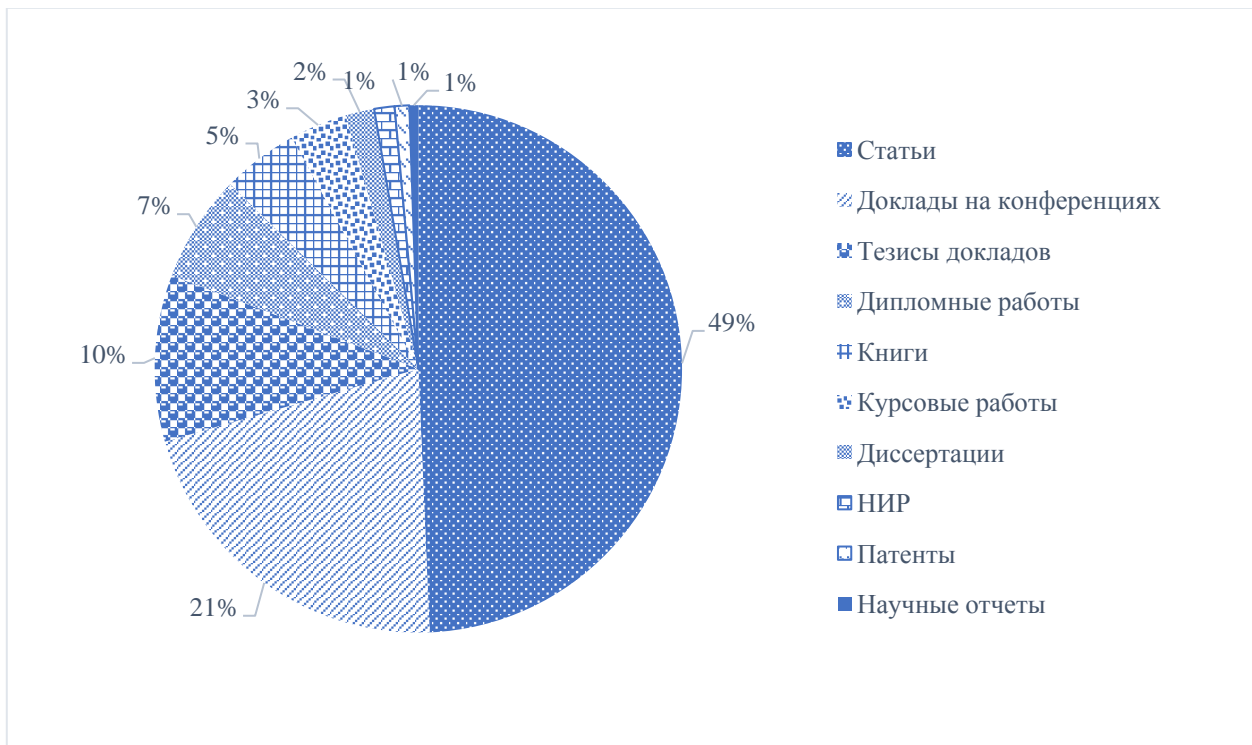


Рис. 1. Типы материалов Интеллектуальной системы тематического исследования наукометрических данных «ИСТИНА»

Однако возникает вопрос видимости собранной и хранящейся на их платформах информации. Ведь именно видимость зарегистрированных данных напрямую зависит от качества их представления на открытых платформах, насколько полно и корректно формируются эти данные.

### МЕТАДАННЫЕ РОССИЙСКИХ ОТКРЫТЫХ ИНСТИТУЦИОНАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ ТЕКУЩИХ ИССЛЕДОВАНИЙ И САЙТОВ

Основные задачи CRIS-систем – это обработка, сбор и обмен метаданными. Автоматизация процесса генерирования информации в CRIS-системах позволяет выполнять перевод неструктурированной информации в структурированную форму [19, 20]. Для определения основного набора метаданных, характерного для отражения информации о результатах научно-исследовательской деятельности, в нашем исследовании данные были собраны и структурированы относительно объекта информации (авторский профиль, публикации, награды, научные проекты) и определены соответствующие этим объектам элементы метаданных.

Анализ российских сайтов и открытых институциональных информационных систем текущих исследований показал, что набор элементов, формирующих информацию о результатах научной деятельности учреждений, существенно различается. На основании анализа представленных метаданных выделено две группы элементов: основные (данные, которые формируют каркас информации о научной деятельности учреждения (базу)) и факультативные (или дополни-

тельные данные, которые обеспечивают исчерпывающую полноту, но реже представлены на страницах сайтов). Метаданные формируемых элементов связаны между собой перекрестными ссылками: из профиля автора к метаданным публикаций, от публикаций, написанных в соавторстве, к профилю другого автора и т. д. При формировании данных о результатах исследований (публикациях) в ИИСТИ, как правило, используется базовый набор элементов Дублинского ядра (Dublin Core, DC), который был разработан в 90-е гг. для улучшения индексирования и повышения видимости создаваемой в сети информации: тип ресурса (type), название (title), создатель (creator), тема (subject), дата (data), источник (source), идентификатор (identifier) и т. д. Однако этот стандарт не позволяет детально описать все объекты информации об исследовательской деятельности научного учреждения, формируемые на сайтах и в ИИСТИ.

**Метаданные авторского профиля.** Профиль автора, формируемый на сайтах научных учреждений и в ИИСТИ, отражает данные, связанные с квалификацией сотрудника, его профессиональной, учебной и научной деятельностью. По частоте размещения основной метаданных авторского профиля являются: ФИО (90,4%), ученая степень (88,5%), звание (88,5%), образование (80,8%), должность (88,5%), отдел или подразделение, в котором работает сотрудник (84,6%), списки публикаций с минимальным набором данных (88,5%). Практически в полном объеме факультативные элементы представлены только в ИИСТИ (Интеллектуальная система тематического исследования наукометрических данных «ИСТИНА» Московского государственного университета им. М.В. Ломоносо-

ва, Исследовательский портал Уральского федерального университета им. первого Президента России Б.Н. Ельцина, Информационно-аналитическая система «Результаты научной деятельности» Астраханского государственного университета и др.) (табл. 2).

Следует отметить, что, в отличие от ИИСТИ, на сайтах научных учреждений списки публикаций в профиле автора представлены не в полном объеме – часто это избранные работы или списки за несколько лет (не всегда актуальные данные за последние 3-5 лет). Учебная деятельность, как правило, обозначается данными о

преподаваемых дисциплинах (75%) и сведениями о повышении квалификации (75%). Кроме того, наблюдается отсутствие единого шаблона в наполнении профиля (у одного сотрудника дана информация об общем стаже работы, у другого – таких сведений нет). Также было отмечено, что на страницах некоторых сайтов научных институтов (Институт молекулярной и клеточной биологии СО РАН, Институт «Международный томографический центр» СО РАН) отдельный профиль есть только у руководителей отделов, а состав сотрудников отдела перечислен в виде списков ФИО.

Таблица 2

**Метаданные открытых российских платформ и институциональных информационных систем текущих исследований**

Категории CRIS	Метаданные	
	Основные элементы	Факультативные элементы
Персональный профиль автора	ФИО, ученая степень, ученое звание, образование, должность (заведующий, научный сотрудник, консультант, доцент, старший преподаватель, куратор проекта и т. д.), отдел / лаборатория, подразделение, в котором работает сотрудник, список публикаций с минимальным набором данных, идентификаторы, достижения	общий стаж, педагогический стаж, стаж научной деятельности, исследовательские интересы ученого, наукометрические показатели публикационной активности, список докладов, участие в научных проектах, участие в работе диссертационных советов, участие в подготовке конференций, членство в редколлегиях научных журналов
Научные проекты	название проекта, руководитель проекта, дата начала / окончания проекта, номер	тип проекта (грант, проект развития, научная группа, совместная лаборатория и т. д.), статус (выполняется, завершено), аннотация, научные концепции, в рамках которых осуществляются исследования, участники (организации, подразделения, сотрудники) проекта, тип источника финансирования (РФФИ, РФФ и т. д.), ключевые слова
Публикации	автор(ы), название публикации, название источника, год издания, том источника, выпуск источника, страницы, DOI	аннотация, язык, дата размещения, статус журнала (BAK, RSCI, Web of Science, Scopus), URL-адрес, ключевые слова, тип публикации (статья, труды конференции, брошюра, монография и т. д.), предметные области, доступ
Награды	тип награды (диплом, благодарственное письмо, премия и т. д.), название награды, автор(ы) / получатель(и) / лауреат, дата присуждения	степень признания (международная, национальная, региональная), аннотация, отдел / подразделение, награждающая организация, страна

Сегодня идентификация очень важна, многие фонды, журналы и издатели просят авторов представлять идентификатор исследователя. Базы данных цитирования, такие как РИНЦ, Scopus, Web of Science, а также Google Scholar, имеют различные типы идентификаторов или профилей исследователей, которые можно использовать для их идентификации. Информационные системы, функционирующие на стандарте CERIF (Common European Research Information Format) позволяют связывать запись об ученом с его уникальными идентификаторами (авторские ID), что способствует повышению видимости автора и его идентификации с результатами научных исследований [6, 9, 21-23]. Соответственно, размещение идентификаторов в профиле автора вполне закономерно – 76,9% российских учреждений на своих платформах уделяют этому внимание. Метаданные при этом формируются в виде значка информационной системы и/или номера идентификатора с гиперссылкой на внешний источник. Такой подход способствует повышению точности и полноты данных, тем самым улучшая их качество.

**Метаданные публикаций.** Для представления данных о результатах исследований (публикациях) базовым является набор основных элементов: автор(ы), название публикации, название источника, год издания, том источника, выпуск источника, страницы, DOI. Следует отметить, что этот набор обязателен для метаданных на сайтах научных учреждений. В соответствии с формированием информации на сайтах и в открытых CRIS-системах остальные элементы метаданных публикаций можно отнести к факультативным: аннотация, язык, дата размещения, статус журнала (BAK, RSCI, Web of Science, Scopus), URL-адрес, ключевые слова, тип публикации (статья, труды конференции, брошюра, монография, патенты, препринты, отчеты и т. д.), предметные области, доступ. Однако формируемое на открытых платформах библиографическое описание в том числе из международных баз данных, не решает проблем, связанных с различным написанием фамилий авторов публикаций (особенности транслитерации), переименованием названий учреждений (влияние реорганизаций), неполными данными, что также отмечено в исследованиях [15, 24]. К проблемам, влияющим на качество метаданных информационных систем, следует отнести также ручной ввод информации сотрудниками организации, что снижает уровень точности представленной информации: орфографические ошибки, многократное и/или неправильное введение данных, несогласованные форматы данных и т. д. [4, 25, 26].

Наиболее полные метаданные публикаций, которые включают в том числе факультативные элементы (хоть и не всегда в полном объеме), как правило, отражены в CRIS-системах. На сайтах университетов и научно-исследовательских институтов публикации представлены в виде общих списков публикаций за последние годы, либо списков публикаций на странице научного сотрудника этого учреждения. Соответственно, метаданные публикаций даются в краткой форме – только основными библиографическими элементами.

Кроме того, следует отметить, что разработчики некоторых ИИСТИ (например, Интеллектуальная система тематического исследования наукометрических данных «ИСТИНА») уделяют большое внимание использованию встроенных библиографических менеджеров (BibTeX, EndNote, RIS, Word, ISI, ADS) для удобства пользователей при оформлении цитирования документа в списке литературы.

**Метаданные о научных проектах,** которые ведутся в учреждении, и **наградах сотрудников** достаточно скромные (см. табл. 2). Основные элементы для отражения информации о проекте – это: название проекта, руководитель проекта, дата начала и окончания проекта, его номер (ЦИТИС, РФФИ, РНФ и т. д.). Для наград и благодарностей сотрудников научных учреждений базовыми метаданными являются такие элементы, как тип диплома (диплом, благодарственное письмо, премия и т. д.), название награды, автор(ы) / получатель(и) / лауреат, дата присуждения. Если данные о научных проектах на сайтах университетов и научно-исследовательских институтов формируются на отдельных страницах, то информация о наградах собрана, как правило, на страницах профиля автора. Ключевым элементом метаданных как для проекта, так и для наград является их название (80,8%).

## ЗАКЛЮЧЕНИЕ

Информационные системы, отражающие текущую научную деятельность, и страницы сайтов университетов и научно-исследовательских институтов являются востребованным источником информации для оценки и организации управления исследовательским процессом. Соответственно формирование на их платформах качественного набора метаданных (полного и корректного) становится важным требованием в процессе администрирования систем.

Анализ показал, что среди собираемых данных, наибольшее внимание уделяется формированию метаданных профилей сотрудников и их публикационной активности. При этом следует отметить недостаточную актуальность и точность представленных на сайтах научных учреждений метаданных: списки публикаций не пополняются работами за последние годы, для некоторых сайтов характерно размещение в профиле сотрудника только избранных работ, а также встречаются ошибки и дублированные записи. Это, в свою очередь, может отрицательно влиять не только на видимость работ авторов в сети и их цитируемость, но и сказываться на показателях результативности самой научной организации.

Кроме того, можно сделать вывод, что использование факультативного набора элементов способствует наиболее полному отражению данных о научно-исследовательской деятельности учреждения. Однако такой набор метаданных формируется только на площадках CRIS-систем, тогда как на страницах сайтов научных учреждений представлены, как правило, только списки с базовым набором элементов. Следует отметить, что аннотация не входит в базовый набор элементов формируемых объектов. Тогда как этот элемент метаданных является очень важным, так как позволяет кратко, в доступной и понят-

ной форме изложить тему исследования, его задачи и результаты, на основании чего можно оперативно сделать соответствующие выводы относительно представленного объекта. Особенно этот элемент важен для формирования метаданных о публикациях и проектах. Для обмена метаданными и их повторного использования важно унифицировать набор элементов для представления результатов научно-исследовательской деятельности в сети.

## СПИСОК ЛИТЕРАТУРЫ

1. Parinov S. Semantic Enrichment of Research Outputs Metadata: New CRIS Facilities for Authors // *Communications in computer and information science*. – 2014. – Vol. 478. – P. 206–217. DOI: 10.1007/978-3-319-13674-5\_20.
2. Зелепухина В.А. Оценка качества открытых данных Роспатента в контексте интеграции с отечественными информационными системами текущих исследований // *Научные и технические библиотеки*. – 2022. – № 12. – С. 15–34.
3. Mugabushaka A.M., Papazoglou T. Information systems of research funding agencies in the “era of the big data”. The case study of the research information system of the European Research Council // *Proceedings of the 11th International Conference on Current Research Information Systems (CRIS 2012)*. – Prague: EuroCRIS, 2012. – P. 103–112.
4. Azeroual O., Schöpfel J. Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries // *Publications*. – 2019. – URL: 10.3390/publications7010014 (дата обращения: 18.06.2025).
5. Chudlarský T., Dvorak J., Dvorak J., Souček M. A Comparison of Research Output Counting Methods Using a National CRIS – Effects at the Institutional Level // *Procedia Computer Science*. – 2014. – Vol. 33. – P. 147–152. DOI: 10.1016/j.procs.2014.06.024.
6. Galimberti P., Mornati S. The Italian model of distributed research information management systems: a case study // *Proceedings of the 13th international conference on current research information systems (CRIS 2016), communicating and measuring research responsibly: profiling, metrics, impact, interoperability*. – 2017. – Vol. 106. – P.183–195.
7. Jetten M., Simons E., Rijnders J. The role of CRIS’s in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands // *Procedia Computer Science*. – 2019. – Vol. 146. – P. 156–165. DOI: 10.1016/j.procs.2019.01.090.
8. Melsheimer B., Marcus W. Introducing CRIS at FAU: Project Presentation // *Proceedings of the 13th international conference on current research information systems (CRIS 2016) communicating and measuring research responsibly: profiling, metrics, impact, interoperability*. – 2017. – Vol. 106. – P. 239–244.
9. Piscanc J. et al. Regional Portal FVG: Effective Interoperability Trough DSpace-CRIS and Open Standards // *Там же*. – P. 82–86.
10. Savic M., Ivanovic M., Radovanovic M., Surla B.D. Towards Culture-Sensitive Extensions of CRISs: Gender-Based Researcher Evaluation // *Model and Data Engineering September. (MEDI 2016) // Lecture Notes in Computer Science*. – 2016. – Vol. 9893. – P. 332–345. DOI: 10.1007/978-3-319-45547-1\_26.
11. Когаловский М.Р., Паринов С.И. Метрики онлайн-информационных пространств // *Экономика и математические методы*. – 2008. Т. 44, № 2. – С. 108–120.
12. Васенин В.А., Афонин С.А., Зензинов А.А., Лунев К.В., Шачнев Д.А. Механизмы системы «ИСТИНА» для интеллектуального анализа состояния и стимулирования хода выполнения проектов в сфере науки и высшего образования // *Научный сервис в сети Интернет*. – 2019. – № 21. – С. 210–221. – URL: 10.20948/abrau-2019-48.
13. Гончаров М.В., Колосов К.А. Формирование расширенных метаданных в информационной системе Единый открытый архив информации ГПНТБ // *Научные и технические библиотеки*. – 2023. – № 1(11). – С. 84-98. DOI: 10.33186/1027-3689-2023-11-84-98.
14. Альперин Б.Л., Ведягин А.А., Зибарева И.В. SciAct – информационно-аналитическая система Института катализа СО РАН для мониторинга и стимулирования научной деятельности // *Труды ГПНТБ СО РАН*. – 2015. – Т. 9. – С. 95–102.
15. Власова С.А., Каленов Н.Е. Многофункциональная веб-система для регистрации и учёта результатов интеллектуальной деятельности учёных // *Научные и технические библиотеки*. – 2022. – № 2. – С. 29–48. DOI: 10.33186/1027-3689-2022-2-29-48.
16. Кабакова Е.А. Автоматизированная система учета результатов интеллектуальной деятельности как инструмент комплексного учета результатов научной деятельности организации (на примере АСУ РИД ВОЛНЦ РАН) // *Вопросы территориального развития*. – 2019. – № 4(49). – С. 7.
17. Шокин Ю.И., Жижимов О.Л., Федотов А.М. Информационные системы ИВТ СО РАН // *Труды XVI Всероссийской конференции «Распределённые информационно-вычислительные ресурсы. Наука – цифровой экономике» – DICR-2017 (г. Новосибирск, 4–7 декабря 2017 г.)*. – Новосибирск: ИВТ СО РАН, 2017. – С. 11–18. – URL: <http://elib.ict.nsc.ru/jspui/bitstream/ICT/1467/5/paper01.pdf> (дата обращения: 22.05.2025).
18. Guskov A. et al. RuCRIS: a pilot CERIF based system to aggregate heterogeneous data of Russian research projects // *Procedia Computer Science*. – 2014. – Vol. 33. – P. 163–167.
19. Бориско С.Н. Информационные системы учета наукометрических данных научных сотрудников // *Материалы 2-й Всероссийской научно-практической конференции «Проблемы повышения эффективности научной работы в оборонно-промышленном комплексе России»*. – Астрахань: Астраханский государственный университет, Издательский дом «Астраханский университет», 2019. – С. 28–35.

20. Vestdam T., Rasmussen H., Doornenbal M. Black Magic Meta Data-get a Glimpse behind the Scene // *Procedia Computer Science*. – 2014. – Vol. 33. – P. 239–244. DOI: 10.1016/j.procs.2014.06.038.
21. Паринов С.И. Международная профессиональная ассоциация разработчиков научных информационных систем euroCRIS и ее главный продукт CERIF // Труды XVI Всероссийской научной конференции (RCDL-2014) «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (г. Дубна, 13-16 октября 2014 г. ) / сост. Л.А. Калмыкова, М.Р. Когаловский. – Дубна: Объединенный институт ядерных исследований, 2014. – С. 26–29.
22. Grenza D. et al. A CRIS in the Desert: The Implementation of Pure at KAUST: A Case Study in Information Exchange // *Procedia Computer Science*. – 2017. – Vol. 106. – P. 176–182. DOI: 10.1016/j.procs.2017.03.014.
23. Dvoráka J., Drobíková B., Bollini A. Publication Metadata in CERIF: Inspiration by FRBR. CRIS // *Procedia Computer Science*. – 2014. – Vol. 33. – P. 47–54. DOI: 10.1016/j.procs.2014.06.008.
24. Васенин В.А., Гаспарянц А.Э. Разрешение неоднозначности имен авторов: анализ публикаций // *Программная инженерия*. – 2017. – Т. 8, № 6. – С. 264–275.
25. Шамаев В.Г., Горшков А.Б. Русскоязычное направление работы российских информационных служб // *Акустический журнал*. – 2020. – Т. 66, № 1. – С. 104–116.
26. Azeroual O., Saake G., Schallehn E. Analyzing data quality issues in research information systems via data profiling // *International Journal of Information Management*. – 2018. – Vol. 41. – P. 50–56. DOI: 10.1016/j.ijinfomgt.2018.02.007.

*Материал поступил в редакцию 11.08.25.*

#### **Сведения об авторе**

**УДАРЦЕВА Ольга Михайловна** – кандидат педагогических наук, старший научный сотрудник Государственной публичной научно-технической библиотеки Сибирского отделения Российской академии наук, г. Новосибирск  
e-mail: udartseva@spsl.nsc.ru

Ю.В. Перепелкина

## О проблемах управления системами электронного документооборота с элементами искусственного интеллекта

*Рассматриваются вопросы внедрения и эксплуатации систем электронного документооборота (СЭДО) на средних и крупных промышленных предприятиях, а также особенности автоматической обработки данных с помощью модулей с искусственным интеллектом. Представлено статистическое исследование мировой тенденции роста СЭДО с модулями искусственного интеллекта. Выделены проблемы использования интеллектуальных автоматизированных процессов обработки документации, заключающиеся, главным образом, в проектировании серверной архитектуры, приложений и машинных методов обработки данных различного типа. Отмечается тенденция интеграции данных кадрового, управленческого и маркетингового типа с данными технологических и производственных процессов.*

**Ключевые слова:** электронный документооборот, бизнес-процессы, оптимизация производства, искусственный интеллект, системы контроля

DOI: 10.36535/0548-0019-2025-11-4

### ВВЕДЕНИЕ

Быстрое развитие информационных технологий и наличие огромного объема данных в современном бизнес-мире вызывают потребность в организации системы электронного документооборота (СЭДО) [1]. Использование искусственного интеллекта (ИИ) для обработки документального потока обеспечивает точность, эффективность и высокую скорость проверки документов, значительно сокращает вероятность ошибок вследствие человеческого и внешних факторов, а также повышает информационную безопасность предприятия [2, 3]. Функционирование СЭДО значительно упрощает процесс автоматизации обработки потоков документов, но, в то же время, приводит к необходимости расширения спектра программных средств модулей с элементами ИИ. Начиная с модулей по распознаванию образов, алгоритмы с элементами ИИ постепенно переходят в область анализа содержимого документов различных типов, включая рукописные подписи и печатные оттиски [3], и связанных с ними данных практически всех блоков бизнес-процесса предприятия.

Высокая степень масштабируемости информационной системы предприятия позволяет обрабатывать огромное количество документов, поддерживая при этом информационную безопасность внешнего и внутреннего потоков документов.

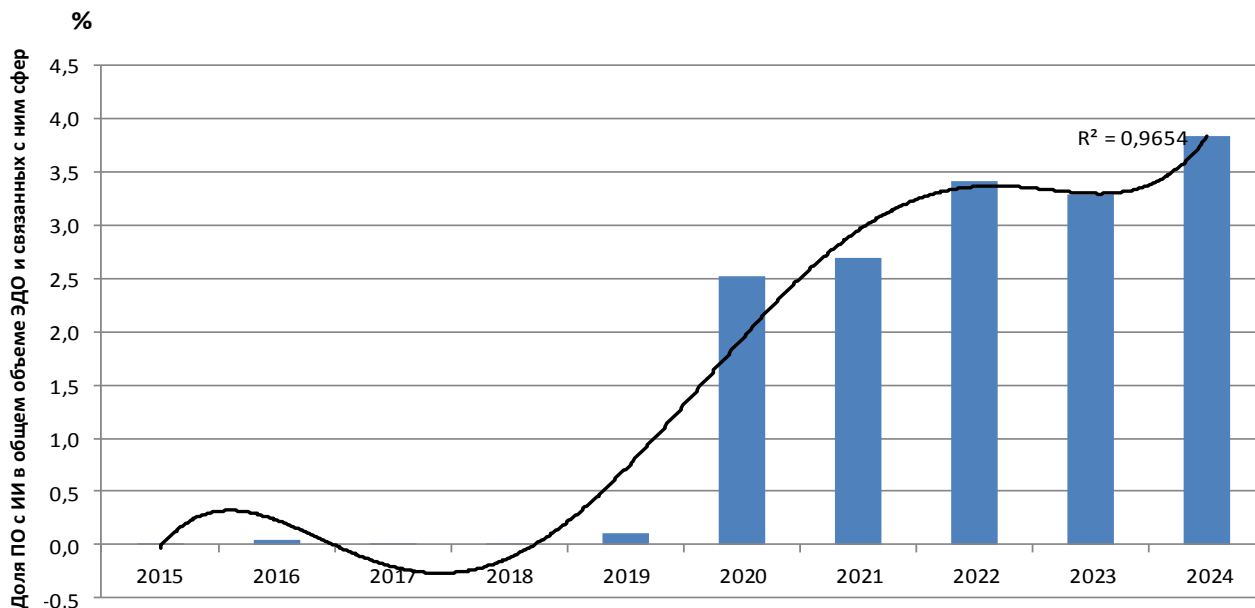
### ЭТАПЫ РЕАЛИЗАЦИИ И ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ СИСТЕМ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

На крупном предприятии СЭДО, как правило, не является единственной ИТ-системой, хотя и незаменима. На промышленных и производственных пред-

приятиях документооборот и технологические процессы тесно переплетены и связывают разные функции в рамках бизнес-процесса [4]. В системе обрабатывается полный ассортимент отчетов по всем видам работ; важное место занимает интеграция с системами информационной безопасности предприятия [5, 6].

За последнее десятилетие объем и разнообразие типов обрабатываемых данных значительно расширились. В целом процесс внедрения интеллектуального программного обеспечения в СЭДО имеет заметно выраженную закономерность, и объемы этого внедрения увеличиваются с каждым годом, охватывая все новые сферы бизнес-процессов. Это происходит из-за того, что существенный объем информационного потока из традиционных областей кадрового обеспечения, финансового планирования и бизнес-аналитики постепенно переходит в области управления и оптимизации процессов и производств, промышленной и производственной инженерии, стандартизации и методов оценки, интернета вещей, больших данных и когнитивных вычислений, анализа производительности, проектирования сетей и инновационных решений, биотехнологий, когнитивных архитектур, коллаборативных и экспертных систем [7].

Анализ результатов проведенного с использованием статистических методов исследования процессов внедрения и функционирования модулей с ИИ в электронный документооборот предприятия различного уровня и масштаба (рисунок) [8, 9], показал, что в целом эти процессы имеют заметно выраженную закономерность (что доказывает высокая точность построенной статистической модели с  $R^2 \approx 0,96$ ), и объемы внедрения увеличиваются с каждым годом,



Эволюция доли модулей с ИИ в общем объеме электронного документооборота предприятий (2015–2024 гг.)

охватывая все новые сферы бизнес-процессов. Алгоритмы машинного обучения активно применяются на промышленных предприятиях, в том числе на тех, где рабочие процессы связаны с физико-химическим производством, многочисленными факторами контроля за состоянием материалов, агрегатных веществ и качеством продукции. Например, цифровым способом реализуются процессы оптимизации складских операций на базе нейросетей [10]. Технологии ИИ на основе решений компании Nvidia, реализованные в сети магазинов Zalando, рассчитывают оптимальные маршруты передвижений работников склада, что с вычислительной точки зрения является одним из частных случаев задачи коммивояжера. Интегрированная в информационную систему и документооборот нейронная сеть оценивает кратчайший маршрут на складе. Работнику отправляются задания на получение нужного количества товара, вводятся промежуточные списки, тем самым рациональнее используется труд работника и повышается общая скорость работы на складе. На некоторых предприятиях внедрена технология визуального распознавания бракованных упаковок и изделий. Предварительно обученная нейросеть идентифицирует визуальные повреждения и геометрические несоответствия стандартам и дает оператору службы контроля качества сигнал перепроверить конкретную единицу товара. Также используются технологии интеллектуального распознавания возникающего брака изделия.

На некоторых предприятиях контроль качества поступающих в производство материалов осуществляется с помощью интегрированного в испытательные установки программного обеспечения с элементами ИИ, включающими разработанные специфические алгоритмы, которые не только снижают трудоемкость

производственного процесса в целом, но и многократно удешевляют его. Для оценки человеческих ресурсов автоматизированные системы электронного документооборота позволяют проводить предварительный подбор персонала, основываясь на процедурах сопоставления навыков сотрудника квалификационным требованиям и уровню сложности рабочего оборудования и процессов.

### ПРОБЛЕМЫ И ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ ИНТЕЛЛЕКТУАЛЬНЫХ ИНТЕГРИРУЕМЫХ ПОДСИСТЕМ

Функционирование крупного промышленного предприятия вызывает высокую нагрузку на его собственную информационную систему. Это требует наращивания аппаратных мощностей и повышения уровня безопасности предприятия в целом.

При этом остается много проблем в области автоматизированной машинной обработки данных, полученных с технологических и производственных объектов. С точки зрения математического аппарата, в обработке данных имеются функциональные и численные ограничения, что затрудняет обобщение методик для решения аналогичных задач. Производители устанавливают датчики на рабочее оборудование для сбора в режиме реального времени информации, которая в обработанном виде отправляется во все подразделения предприятия для обеспечения взаимодействия структурных подразделений и принятия соответствующих управленческих решений. Оптимизация, контроль и прогнозирование на разных этапах бизнес-процессов имеют огромное значение, особенно на тех предприятиях, бизнес-процессы которых связаны с повышенной опасностью (химическое производство, атомные установки, системы экологи-

ческого контроля и жизнеобеспечения), дорогостоящими материалами и сложным в эксплуатации оборудованием [9].

Корпоративные информационные системы с объемной и сложной структурой, обладая большим количеством пользователей и широкой территориальной их распределённостью, имеют проблемы с администрированием. Для оптимизации бизнес-процессов таких информационных систем рекомендуется реализовывать управление ими из единого центра с минимальными трудозатратами на обслуживание.

Сегодня использование систем со встроенным ИИ для проверки и обработки документов различного типа в компании и особенно в корпорации с огромными объемами документооборота стало устойчивой тенденцией развития цифровизации современного бизнеса.

## ЗАКЛЮЧЕНИЕ

Обобщая результаты проведенного анализа и учитывая состояние и эволюцию современных систем электронного документооборота, можно сделать вывод о том, что для их эффективного функционирования и перспективного развития необходимы следующие условия:

1) СЭДО должна быть интегрирована с информационными системами и потоками компании, включая процессы обработки справочной информации, менеджмента и передачи данных в режиме реального времени при сквозных бизнес-процессах;

2) различные структура и внешний вид интерфейса СЭДО должны отвечать функциональным потребностям пользователей разных категорий и навыков, а также интенсивности обработки данных на крупных предприятиях, где количество сотрудников, как правило, достигает нескольких тысяч;

3) для корпоративных ИТ-систем с высокой информационной нагрузкой при установке программного обеспечения больших объемов рекомендуется предусмотреть автоматическую (без привлечения администратора) инсталляцию и/или обновление модулей СЭДО с сервера.

Репрезентативные примеры использования гибридных платформ с элементами ИИ показывают, что применение интеллектуальных методов в процессе автоматизированной обработки документооборота имеет огромный потенциал для оптимизации бизнес-процессов и повышения эффективности работы предприятия в целом.

## СПИСОК ЛИТЕРАТУРЫ

1. СЭД Docsvision. – URL: <https://docsvision.com> (дата обращения 23.08.2025).

2. Лысачев М.Н., Прохоров А.Н. Искусственный интеллект. Анализ, тренды, мировой опыт / науч. ред. Д.А. Ларионов. – Корпоративное издание. – Москва; Белгород: КОНСТАНТА-принт, 2023. – 460 с.:
3. Компания Bellerage. – URL: <https://www.acg.ru/ru/services/it-services> (дата обращения 23.08.2025).
4. Перепелкина Ю.В., Гришин А.А. Тенденции внедрения систем искусственного интеллекта в экспериментальные исследования для целей промышленности и производства // Сб. трудов XXXIV Международной научно-практической конференции «Предотвращение. Спасение. Помощь». – Химки, Моск. обл.: Из-во Академии гражданской защиты МЧС России им. генерал-лейтенанта Д.И. Михайлика, 2024. – С. 95-104.
5. +95 Cyber Security Breach Statistics 2025. – URL: <https://arxiv.org/abs/2308.02479> (дата обращения: 23.08.2025).
6. Zyryanov S.I., Berezhnov V.P., Perepelkina Y.V., Shevtsov A.I., Shevchenko K.K., Kozlov V.V. The production model of fuzzy neural network in information security systems // International Journal of Advanced Trends in Computer Science and Engineering. – 2020. – Vol. 9, № 4. – P. 5190-5192. – URL: <https://repository.rudn.ru/recordsources/recordsource/10124> (дата обращения: 23.08.2025).
7. Robotics: Science and System XIX // Republic of Korea conference web-site, July 10 – July 14, 2023 / eds. by K. Bekris, Kr. Hauser, S. Herbert, J. Yu. – 2023. – URL: <https://www.roboticsproceedings.org/rss19/index.html> (дата обращения: 23.08.2025).
8. TAdviser Statistics: 2025. – URL: <https://www.tadviser.ru> (дата обращения 23.08.2025).
9. Scimago Journal & Country Ranks 2015-2024. – URL: <https://www.scimagojr.com> (дата обращения 23.08.2025).
10. Forbes. – URL: <https://www.forbes.ru/tehnologii/340635-robot-na-razborke-pochemu-sklad-bez-lyudey-segodnya-na-samom-dele-nikom-ne-nuzhen> (дата обращения 23.08.2025).

*Материал поступил в редакцию 29.08.25.*

## Сведения об авторе

**ПЕРЕПЕЛКИНА Юлианна Вячеславовна** – кандидат физико-математических наук, (1) доцент кафедры инфокогнитивных систем Московского политехнического университета; зав. Отделом научных исследований по механике ВИНТИ РАН, Москва e-mail: amadeycity@yandex.ru

# ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК 016:004.6.057.8

Н.С. Солошенко, О.В. Федорец

## Формирование междисциплинарных тематических коллекций документов с использованием библиографических баз данных открытого доступа\*

*Исследованы возможности использования библиографических баз данных открытого доступа (The Lens, Dimensions, OpenAlex) и открытой программы визуализации данных VOSViewer для формирования междисциплинарных тематических коллекций документов и составления перечней наиболее продуктивных и релевантных научных журналов по отдельным приоритетным направлениям научно-технологического развития. Отмечена неполнота метаданных у многих документов (отсутствие рефератов и ключевых слов) в перечисленных базах данных. Представлен опыт формирования и статистического анализа тематических коллекций статей и перечней журналов по композитной керамике. Изучено рассеяние профильных статей по журналам, а также дана количественная оценка сходства ядерных зон журналов, выявленных в 2018 и 2024 гг. Дана сравнительная характеристика исследуемых баз данных по нескольким критериям.*

**Ключевые слова:** библиографические базы данных, тематические коллекции документов, тематические перечни журналов, рассеяние научных статей, коэффициент сходства Жаккара, перспективная композитная керамика

DOI: 10.36535/0548-0019-2025-11-5

### ВВЕДЕНИЕ

Актуализация комплектования крупного информационного центра научно-технической литературой (НТЛ) требует оптимизации отбора информационных ресурсов и разработки эффективного инструментария для исследования мирового потока по приоритетным направлениям развития науки и технологий, определенным Указом Президента РФ "Об утверждении приоритетных направлений научно-технологического развития и перечня важнейших наукоемких технологий"<sup>1</sup>.

В связи с введением санкций в отношении России со стороны недружественных государств после 2022 г.,

уходом с российского рынка проприетарных информационно-аналитических баз данных: Web of Science (Clarivate, США) и Scopus (Elsevier, Нидерланды) и изменением положения России в мировом информационном пространстве, существенно вырос интерес российских исследователей и научных администраторов к международным информационно-библиографическим ресурсам открытого доступа. В профильной литературе появились многочисленные публикации, в которых авторы рассматривают функциональные возможности библиографических БД открытого доступа для решения различных задач. Так, в работе А.В. Лутая и Е.Э. Любушко [1] приведены результаты детального анализа индексируемых этими системами метаданных. Авторы обращают внимание пользователей на то, что в базах данных открытого доступа, в частности The Lens, Semantic Scholar и OpenAlex, используются открытые метаданные, полученные из Medline, Microsoft Academic<sup>2</sup> и CrossRef, и отмечают, что зависимость от одинако-

\* Работа выполнена в рамках государственного задания ВИНТИ РАН по теме № FFFU-2025-0009 «Аналитические исследования динамики трансформации предметных профилей сериальных изданий по приоритетным направлениям науки, техники и технологий».

<sup>1</sup> Указ Президента РФ от 18.06.2024 № 529 "Об утверждении приоритетных направлений научно-технологического развития и перечня важнейших наукоемких технологий". – 2024 – URL: <http://publication.pravo.gov.ru/document/0001202406180018> (дата обращения: 01.07.2025)

<sup>2</sup> Проект Microsoft Academic закрыт в конце 2021 г.

вых источников может стать причиной повторения в разных базах данных схожего набора пустых полей и ошибок. Вместе с тем результаты некоторых исследований [2, 3] указывают на то, что в открытых БД заимствованные метаданные обогащаются дополнительной информацией за счет добавления реферата, типа документа, статуса открытости и др. В публикациях [1, 4–7] сравниваются характеристики основных открытых международных политематических библиографических ресурсов, возникших в последние десятилетия (в том числе: AMiner, Google Scholar, Semantic Scholar, открытой версии Dimensions, Lens, OpenAlex), с аналогичными характеристиками коммерческих систем. Авторы показывают, что состав источников у большинства недавно образованных информационных систем существенно выше, чем у Web of Science и Scopus, что подтверждается, например, исследованиями V.K. Singh et al., определившими, что в Dimensions индексируется на 82,22% больше журналов, чем в Web of Science и на 48,7% больше, чем в Scopus [8].

В.Н. Гуреев и Н.А. Мазов в своем обзоре [4] отметили такие важные критерии при выборе баз данных для поисковой работы и библиометрических исследований, как полнота состава метаданных и полнота их выгрузки с возможностью выбора форматов, а также широкий набор фильтров при поиске, реализация расширенного поиска, создание глобальных и сводных запросов и др. Авторы представили степень соответствия анализируемых систем приведенным критериям и аргументировали вывод о возможности использования БД The Lens и открытой версии Dimensions как практически полноценной замены коммерческих продуктов.

Широкое наполнение и активное развитие сервисов открытых библиографических ресурсов позволяют рассматривать их как альтернативу при решении многих информационно-поисковых задач, например: использование открытых ресурсов на разных стадиях производства научного знания [6], работа с публикационными профилями авторов и научных организаций [9], изучение трендов совместной публикационной активности авторов [10], определение фронтиров науки [11].

Цель настоящей работы – исследование возможностей библиографических БД открытого доступа для формирования междисциплинарных тематических коллекций документов и составления перечней наиболее продуктивных и релевантных сериальных изданий (СИ) по отдельным приоритетным направлениям научно-технологического развития с использованием открытой программы визуализации данных VOSViewer [12].

Важнейшими наукоемкими сквозными технологиями, перечисленными в вышеупомянутом Указе Президента РФ, являются «технологии создания новых материалов с заданными свойствами и эксплуатационными характеристиками». Это – междисциплинарные технологии, и они включают множество тематик, поэтому их невозможно выразить точным набором классификационных индексов. Тематические профи-

ли коллекций документов по каждой из тематик наиболее точно описываются перечнями ключевых слов, которые используются в качестве предметных рубрик предметных профилей [13].

Одним из направлений в области разработки новых перспективных материалов является создание высокоэнтропийных материалов, состоящих, по крайней мере, из пяти элементов или компонентов и демонстрирующих в качестве основных характеристик: высокую энтропийность, сильное искажение кристаллической решетки, слабую диффузию и коктейль-эффекты, которыми обычно объясняют механизмы возникновения различных феноменов, связанных с этой категорией многокомпонентных систем [14]. Информация об этих материалах появилась в 2004 г., когда были опубликованы первые результаты исследований по высокоэнтропийным сплавам с многообещающими свойствами [15]. В дальнейшем научные разработки стали проводиться на различных многокомпонентных системах, в том числе и на композитной керамике, что привело к созданию термостойких, изностостойких и антикоррозионных покрытий, термоэлектриков, катализаторов и многофункциональных пористых материалов [16].

## МЕТОДИКА И ИНСТРУМЕНТАРИЙ

С целью изучения возможности использования библиографических БД открытого доступа для формирования междисциплинарных тематических коллекций была выбрана тематика в области новых материалов – перспективная композитная керамика (ПКК).

Публикации по этой тематике рассеяны в изданиях по многим научным областям, поэтому собрать тематические коллекции возможно только с использованием ключевых слов в поисковых запросах.

Для отбора статей в междисциплинарные тематические коллекции по ПКК были составлены поисковые запросы в открытые библиографические ресурсы:

- The Lens (Австралия) – 287,1 млн научных публикаций;
- Dimensions (Великобритания), бесплатная версия – 155,4 млн публикаций;
- OpenAlex – более 250 млн научных публикаций.

Отбор статей выполнялся с учетом правил поиска в этих ресурсах в следующих метаданных:

- БД The Lens – Title, Abstract, Keyword or Field of Study;
- БД Dimensions – Title, Abstract;
- OpenAlex: Title, Abstract.

В табл. 1 приведены поисковые запросы по перспективной композитной керамике в эти БД.

Тематические коллекции статей по ПКК в БД The Lens формировались по результатам запросов с использованием двух различных классификаций:

- 1) предметных категорий (Subjects),
- 2) областей исследования (Fields of Study).

В работе использовался также инструментарий программы VOSViewer, Version 1.6.19, позволяющий анализировать как библиографическую, так и текстовую информацию [12].

**Поисковые запросы для отбора статей в тематические коллекции  
по перспективной композитной керамике**

БД The Lens	БД Dimensions	OpenAlex
ceramic* AND (composite* OR poros* OR porous* OR membrane* OR "high entropy" OR "ultra high temperature")  Year Published = (2018- 2024) Publication Type = ( journal article )  1. Subjects <sup>1</sup> = (Materials Chemistry, Ceramics and Composites, General Materials Science, Electronic, Optical and Magnetic Materials; Surfaces, Coatings and Films; Mechanics of Materials; Process Chemistry and Technology; General Chemistry)  2. Fields of Study <sup>2</sup> : Materials science ; Ceramic, Composite material, Porosity, Chemical engineering, Microstructure, Membrane, Coating, Chemistry, Dielectric, Ceramic matrix composite, Ceramic membrane, Nanocomposite, Spark plasma sintering, Oxide,	ceramic AND (composite OR poros OR porous OR membrane OR "high entropy" OR "ultra high temperature")  Publication Year: 2018 OR 2019 OR 2020 OR 2021 OR 2022 OR 2023 OR 2024 Publication Type: Article  Fields of Research (ANZSRC 2020) <sup>3</sup> : 4016 Materials Engineering 34 Chemical Sciences 4004 Chemical Engineering	ceramic* AND (composite* OR poros* OR porous* OR membrane* OR "high entropy" OR "ultra high temperature")  Type: article  Year: 2018, 2019, 2020, 2021, 2022, 2023, 2024

**ПРИМЕЧАНИЯ:**

1. Предметные рубрики (Subjects) извлекаются из описаний журналов в метаданных из Crossref.
2. Области исследования (Fields of Study) определяются на основе машинного обучения и парсинга всех доступных текстовых данных в описаниях статей.
3. Области исследования (Fields of Research) Australian and New Zealand Standard Research Classification (ANZSRC).

### **ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ИНСТРУМЕНТАРИЕВ БД ОТКРЫТОГО ДОСТУПА**

В БД The Lens по результатам запроса возможно выделить не более 100 источников, ранжированных по убыванию профильных публикаций. Вместе с тем, обработка массива метаданных, извлеченных по результатам запроса в БД, программой VOSViewer позволяет получить полный перечень сериальных изданий с показателем цитирования статей.

В БД Dimensions при формировании запроса предоставляется возможность выбрать интересующие области исследования (Fields of Research) из перечня, включающего 100 наименований. В этой БД также возможно выделить перечень продуктивных журналов с использованием опции Analytical Views / Source Titles.

Выгрузка метаданных из БД Dimensions производится без авторских ключевых слов. Максимальное количество документов в одной выгрузке составляет 2500.

Инструментарий БД OpenAlex предоставляет возможность одновременно сформировать тематическую коллекцию только по одной из научных областей (fields) или предметных категорий (subfields), поэтому коллекция по ПКК из этой БД формировалась без ограничений по научным областям.

В OpenAlex невозможно выделять перечни журналов без использования API (Application Programming Interface) – интерфейса прикладного программирования, позволяющего различным программным приложениям взаимодействовать друг с другом. Поэтому в исследовании по формированию ранжированных перечней этот ресурс не применялся.

Для оценки полноты наличия ключевых слов (КС) использовались массивы метаданных публикаций, включенных в тематические коллекции по ПКК в БД The Lens и OpenAlex, поскольку выгрузка ключевых слов из открытой версии БД Dimensions не предусмотрена.

С использованием предлагаемых аналитических инструментариев в БД The Lens и Dimensions были выделены перечни сериальных изданий (СИ), включающих статьи по результатам запросов в каждый заданный хронологический период.

В качестве меры пересечения двух перечней сериальных изданий, использовался коэффициент сходства Жаккара:

$$K_J = c/(a+b-c), \quad (1)$$

где: a – количество журналов в 1-м списке;  
b – количество журналов во 2-м списке;  
c – количество журналов, общих для 1-го и 2-го списков.

## ФОРМИРОВАНИЕ ТЕМАТИЧЕСКИХ КОЛЛЕКЦИЙ СТАТЕЙ ПО ПКК

В тематические коллекции по ПКК включены релевантные запросам статьи, отражённые в БД The Lens, Dimensions и OpenAlex в 2018-2024 гг.

Содержание и полноту тематической выборки статей определяют:

- состав профильных источников;
- набор классификационных рубрик;
- полнота загрузки выпусков профильных источников в БД;
- полнота метаданных (наличия рефератов, авторских ключевых слов).

Тематические коллекции статей по ПКК формировались по результатам запросов с использованием различных классификаций в БД:

- The Lens – по совокупности определенных предметных категорий (Subjects) и по совокупности определенных областей исследования (Fields of Study);
- Dimensions – по совокупности определенных областей исследования (Fields of Research) Australian and New Zealand Standard Research Classification (ANZSRC).

В БД OpenAlex возможно формировать тематические выборки по 26 различным научным областям (fields) или / и 200 предметным категориям (subfields). Но, так как одновременно можно выбрать только одну научную область или предметную категорию, то вы-

борка статей по ПКК в БД OpenAlex была сформирована без ограничения определенной тематической областью.

Таким образом, с использованием аналитических инструментариев этих ресурсов была получена статистика распределения профильных статей по годам (табл. 2) и перечни наиболее продуктивных по тематике ПКК журналов в 2018 и 2024 гг.

Данные в табл. 2 демонстрируют значительные различия в объемах тематических коллекций журнальных статей по ПКК в анализируемых ресурсах. Тематическая выборка с наибольшим количеством профильных статей была сформирована по результатам запроса в БД The Lens по совокупности нескольких областей исследования (Fields of Study), которые определяются на основе машинного обучения и парсинга всех доступных текстовых данных в описаниях статей. Тематическая выборка в этой БД, сформированная по совокупности выбранных предметных категорий (Subjects), которые извлекаются из метаданных описаний журналов в Crossref, содержит в 2,7 раза меньше (15497 / 41996) документов. Объемы тематических коллекций в БД Dimensions и OpenAlex по количеству статей ближе по значениям к объемам коллекций из БД The Lens, сформированной по набору предметных категорий.

Полнота тематической выборки в БД также зависит от полноты метаданных, в частности, наличия рефератов (табл. 3) и ключевых терминов (табл. 4).

Таблица 2

**Статистика распределения по годам количества статей в тематической коллекции по ПКК в БД открытого доступа\***

Год	The Lens		Dimensions	OpenAlex
	Subjects	Fields of Study		
2018	2006	3717	2272	2363
2019	2153	4182	2621	2516
2020	2297	4681	2788	2576
2021	2706	6338	3163	2713
2022	2115	10050	3430	2877
2023	2068	7685	3620	2905
2024	2152	5343	4103	2905
Всего	15497	41996	21997	18855

*ПРИМЕЧАНИЕ.* \*Количество документов, выгруженных из БД на 25.07.2025 г.

Таблица 3

**Доля метаданных с наличием рефератов в БД The Lens и OpenAlex\***

Количество статей	The Lens (Subjects)		The Lens (Fields of Study)		OpenAlex	
	2018	2024	2018	2024	2018	2024
Всего	2002	2152	3717	5343	2363	2905
С рефератом	1892	1481	3602	4456	1818	2141
Доля статей с рефератом,%	94,5	68,8	96,9	83,4	76,9	73,7

*ПРИМЕЧАНИЕ.* \*В БД Dimensions не представлена информация о доле метаданных статей с рефератами

## Наличие ключевых слов в метаданных статей тематических коллекций по ПКК в различных БД

Количество	The Lens (Subjects)		The Lens (Fields of Study)		OpenAlex		ScienceDirect	
	2018	2024	2018	2024	2018	2024	2018	2024
Документов в выборке	2002	2150	3717	5343	2363	2905	605	1069
Ключевых слов (КС), частота $\geq 1$	919	1389	1187	945	1297	1214	1833	3104
КС / документов	0,46	0,65	0,32	0,18	0,55	0,42	3,03	2,90

По данным, представленным в системах The Lens и OpenAlex, наибольшая доля метаданных статей 2024 г. по ПКК с рефератами выявлена в тематической коллекции The Lens (Fields of Study) – 83,4%, а наименьшая в The Lens (Subjects) – 68,8%. Во всех трех коллекциях эта доля уменьшилась в 2024 г. по сравнению с 2018 г.

Тематические коллекции метаданных статей по ПКК 2018 и 2024 гг. в ресурсах: The Lens и OpenAlex были загружены в программу визуализации библиографических данных VOSViewer для выделения ключевых слов и словосочетаний. К сожалению, из БД Dimensions метаданные статей в тематических выборках выгружаются без ключевых слов.

Полученные результаты по наличию ключевых слов в метаданных статей в различных БД приведены в табл. 4.

Для сравнения полноты включения ключевых слов в метаданные статей, размещенных в БД открытого доступа и на издательских платформах, был выполнен аналогичный запрос (см. табл. 1) на платформе ScienceDirect и выявлены профильные статьи по тематике ПКК в трех высокопродуктивных журналах: *Ceramics International*; *Journal of the European Ceramic Society*; *Journal of Alloys and Compounds* (см. табл. 4). Тематические коллекции метаданных статей по ПКК 2018 и 2024 гг. на платформе ScienceDirect также были загружены в программу визуализации библиографических данных VOSViewer для выделения ключевых терминов. Выявлено, что количество ключевых слов, приходящихся на один документ в тематической коллекции ScienceDirect, на порядок выше, чем в БД открытого доступа.

Меньше всего ключевых слов на один документ приходится в The Lens (Fields of Study): 0,32 КС в 2018 г. и 0,18 КС в 2024 г.

### РАНЖИРОВАННЫЕ ПЕРЕЧНИ СЕРИАЛЬНЫХ ИЗДАНИЙ, ВКЛЮЧЕННЫХ В ТЕМАТИЧЕСКИЕ КОЛЛЕКЦИИ ПО ПКК

В тематических коллекциях The Lens (Fields of Study) и The Lens (Subjects) по результатам запросов было выделено максимально возможное количество источников – по 100 сериальных изданий (СИ), ранжированных в порядке убывания тематической продуктивности (ТПр). Тематическая продуктивность определяется как количество документов из данного источника, относящихся к выбранной тематической коллекции статей по ПКК. По аналогии было выяв-

лено 100 наиболее продуктивных по ПКК сериальных изданий в БД Dimensions.

В случае ранжирования СИ по убыванию тематической продуктивности предполагается рассеяние профильных публикаций по источникам в соответствии с эмпирической формулой С. Брэдфорда:

$$P : P_1 : P_2 = 1 : n : n^2, \quad (2)$$

где:  $P$  – число журналов в первой зоне;  
 $P_1$  – число журналов во второй зоне;  
 $P_2$  – число журналов в третьей зоне;  
 $n$  – коэффициент пропорциональности или рассеивания и количество статей в зонах примерно равное.

В соответствии с наукометрической моделью Брэдфорда в оба периода (2018 и 2024 гг.) были выделены три зоны, содержащие примерно равное количество статей в тематических коллекциях The Lens (Subjects), The Lens (Fields of Study) и Dimensions, и выявлены журналы, входящие в эти зоны.

В табл. 5 представлены данные по степени покрытия зон Брэдфорда журналами, включенными в TOP 100 сериальных изданий в обеих БД.

Таким образом, в БД The Lens и Dimensions были выявлены журналы, покрывающие первые две зоны по Брэдфорду (ядерные и профильные источники), включающие  $\approx 2/3$  статей в 2018 и 2024 гг., при том, что перечни 2018 и 2024 годов содержат по 100 журналов.

Для определения доли пересечения двух перечней сериальных изданий, актуальных в разные статистические периоды (2018 и 2024 гг.), использовался коэффициент сходства Жаккара (1).

Перечень наиболее продуктивных сериальных изданий, входящих в тематическую коллекцию ПКК в оба статистических периода в БД The Lens (Subjects), включает 68 названий СИ. Коэффициент сходства Жаккара для перечней СИ 2018 и 2024 годов рассчитывался как:

$$K_{J_{Lens}} = 68 / (100 + 100 - 68) = 51,52\%.$$

Этот перечень включает источники с высоким уровнем авторитетности: 90% СИ публикуются крупнейшими издателями – Elsevier, Wiley, Springer, MDPI, и в него входит переводная версия российского журнала "Стекло и керамика" (*Glass and Ceramics*) издательства Springer.

**Распределение TOP 100 наиболее продуктивных сериальных изданий по ПКК по зонам Брэдфорда**

Количество статей	The Lens (Subjects)		The Lens (Fields of Study)		Dimensions	
	2018	2024	2018	2024	2018	2024
Всего	2002	2152	3717	5343	2273	4103
TOP 100 СИ	1703	1851	2546	3035	1556	2667
Доля статей в TOP 100 СИ, %	85	86	68,5	56,8	68	65
1 зона СИ (ядро)	659 (4 СИ)	740 (5 СИ)	1249 (8 СИ)	1816 (15 СИ)	751 (12 СИ)	1336 (11 СИ)
2 зона СИ (профильные)	680 (31 СИ)	707 (30 СИ)	1255 (85 СИ)	1219 (85 СИ)	753 (76 СИ)	1331 (89 СИ)
3 зона СИ (факультативные)	364 (65 СИ)	404 (65 СИ)	42 (7 СИ)		52 (12 СИ)	

Таблица 6

**Наиболее продуктивные сериальные издания по ПКК в БД Dimensions и The Lens (Subjects) в 2024 г.**

№ п/п	Название журнала	Dimensions			The Lens (Subjects)			ISSN	Страна издания	Издатель
		ТПр	Зона СИ	Цит. ст.	ТПр	Зона СИ	Цит. ст.			
1	Journal of the European Ceramic Society	181	1	6,9	103	1	6,85	0955-2219	Нидерланды	Elsevier
2	Journal of Alloys and Compounds	101	1	6,74	39	2	6,64	0925-8388	Нидерланды	Elsevier
3	ACS Applied Materials & Interfaces	51	1	5,12	60	2	4,77	1944-8244	США	ACS
4	Ceramics International	555	1	4,26	325	1	4,75	0272-8842	Великобритания	Elsevier
5	Journal of the American Ceramic Society	85	1	2,66	107	1	2,32	0002-7820	США	Wiley
6	Materials	68	1	2,37	109	1	2,26	1996-1944	Швейцария	Elsevier
7	Journal of Materials Science: Materials in Electronics	43	1	2,23	52	2	1,71	0957-4522	США	MDPI
8	International Journal of Applied Ceramic Technology	64	1	1,33	96	1	1,23	1546-542X	США	Science Press

Перечень наиболее продуктивных сериальных изданий, составляющих тематическую коллекцию ПКК в оба статистических периода в БД The Lens (Fields of Study), включает 47 названий журналов. Коэффициент сходства Жаккара для перечней СИ 2018 и 2024 годов рассчитывался как:

$$K_{J \text{ Lens (Fields of Study)}} = 47 / (100+100-47) = 30,72\%.$$

75% журналов этого перечня публикуются крупными издателями – Elsevier, MDPI, Springer, Wiley, а также содержат ряд переводных версий российских журналов: "Стекло и керамика" (Glass and Ceramics); "Неорганические материалы" (Inorganic Materials); "Новые огнеупоры" (New Refractories). Вместе с тем в этот перечень входят и ресурсы открытого доступа, публикующие труды конференций разного уровня (E3S Web of Conferences). Ядерная зона перечня 2018 г. также содержит открытый интернет-ресурс трудов конференций (MATEC Web of Conferences), а в ядерную зону 2024 г. входят три журнала на китайском

языке: Gongchengxue Yanjiu yu Shiyong = Engineering Science Research and Application (ISSN 2717-5316); Gongcheng Guanli yu Jishu Tantaos = Engineering Management and Technical Research, (ISSN 2717-5189); Chengshi Jianzhu yu Fazhan = Urban Architecture and Development (ISSN 2717-557X), в настоящее время не включенные в международные индексы научного цитирования.

Перечень наиболее продуктивных сериальных изданий, входящих в тематическую коллекцию ПКК в оба статистических периода в БД Dimensions, включает 56 названий. Коэффициент сходства Жаккара для перечней СИ 2018 и 2024 гг. рассчитывался как:

$$K_{J \text{ Dimensions}} = 56 / (100+100-56) = 38,89\%.$$

Таким образом, состав источников в тематической коллекции The Lens (Fields of Study) оказался более динамичен во времени (30,72%), чем в коллекциях Dimensions и The Lens (Subjects), где коэффициенты Жаккара были равны, соответственно, 38,89% и 51,52%.

В табл. 6 приведены данные о наиболее продуктивных по ПКК журналах 1-й зоны Брэдфорда, общих для БД Dimensions и The Lens (Subjects) в 2024 г. Журналы ранжируются в порядке уменьшения средней цитируемости статей 2024 г. в БД Dimensions.

Следует отметить, что аналитический инструментарий БД Dimensions позволяет получать сведения о средней цитируемости статей в конкретный период. Хотя инструментарий БД The Lens такой возможности не предоставляет, при обработке массива метаданных БД The Lens программой VOSViewer возможно выделить список источников за соответствующий статистический период с показателем количества цитирований, полученных профильными статьями из этого источника, а также вычислить среднюю цитируемость статей (см. табл. 6). Показатели средней цитируемости профильных статей из сериальных изданий, общих для БД Dimensions и The Lens, различаются незначительно.

Таким образом, в выборке источников для БД Dimensions и The Lens возможно выявлять журналы не только с высокой продуктивностью, но и с высокой средней цитируемостью статей.

Данные табл. 6 демонстрируют высокий уровень авторитетности включенных в коллекцию БД Dimensions и The Lens (Subjects) источников по ПКК. Так, 100% СИ ядерной зоны публикуются крупнейшими издателями: Elsevier, Wiley, Springer, American Chemical Society, MDPI.

Вместе с тем, приведенные данные демонстрируют значительное различие в количестве профильных статей в выборках из одних и тех же источников в БД The Lens и Dimensions.

Например, тематическая продуктивность (ТПр) журнала Ceramics International в 2024 г. составила: в Dimensions – 555 статей, а в The Lens (Subjects) – 325.

Этот же показатель журнала Journal of Alloys and Compounds в Dimensions в 2018г. в 2,6 раза превосходил аналогичный показатель в The Lens – (101 / 39 статей).

Существенная разница в наполнении тематических выборок статьями из конкретных источников в базах данных может быть связана, прежде всего, с различной полнотой представления метаданных. Так, несмотря на более широкий набор полей для поиска профильных публикаций в The Lens (Title, Abstract, Keyword or Field of Study) по сравнению с Dimensions (Title, Abstract), тематическая выборка The Lens (Subjects) в 1,4 раза меньше, чем в Dimensions (15497 / 21997).

Как уже было отмечено, только 68,8% документов из профильной выборки The Lens (Subjects) в 2024 г. содержат рефераты, а в метаданных тематической коллекции The Lens (Fields of Study) выявлена очень невысокая полнота авторских ключевых слов (0,32 КС приходится на один документ в 2018 г. и 0,18 КС в 2024 г.). Эта ситуация объясняется, по-видимому, наличием в ядерной зоне источников The Lens (Fields of Study) открытых трудов конференций, метаданные которых не содержат, в основном, авторские ключевые термины. Неполнота метаданных снижает достоверность статистических исследований.

## ЗАКЛЮЧЕНИЕ

В настоящем исследовании проанализированы возможности использования зарубежных информационных ресурсов открытого доступа: The Lens, OpenAlex и Dimensions (открытая версия) для формирования тематических коллекций по одному из приоритетных направлений научно-технологического развития на примере перспективной композитной керамики (ПКК). В табл. 7 приведены результаты сравнения.

Таблица 7

**Возможности использования открытых библиографических ресурсов для формирования тематических коллекций**

Характеристики БД	The Lens	Dimensions (открытая версия)	OpenAlex
Расширенный запрос	+	+	+
Набор фильтров	+	+	+
Выбор предметных областей	области исследования (Fields of Study) и предметные рубрики (Subjects)	области исследования (Fields of Research) Australian and New Zealand Standard Research Classification (ANZSRC)	единовременно выбор только одной из научных областей (fields) или предметных категорий (subfields)
Возможность выгрузки метаданных	+	max 2500 документов	+
Статистика количества статей по годам	+	+	+
Наличие рефератов	2024 г.: ≈69 % (Subjects); ≈83 % (Fields of Study),	информация отсутствует	2024 г.: ≈74%
Перечень источников	100 наиболее релевантных*	+**	только по API

### ПРИМЕЧАНИЯ:

\* При обработке массива метаданных программой VOSViewer возможно выделить список источников в соответствующий статистический период с показателем количества цитирований, полученных профильными статьями из этого источника.

\*\* Представлена информация о средней цитируемости статьи.

В тематических коллекциях The Lens (Subjects), The Lens (Fields of Study) и Dimensions выделены перечни наиболее продуктивных сериальных изданий 2018 и 2024 годов по тематике ПКК. Перечни The Lens (Subjects) и Dimensions в основном содержат журналы крупнейших авторитетных издателей: Elsevier, Wiley, Springer, American Chemical Society, MDPI. Перечни The Lens (Fields of Study) помимо авторитетных журналов включают открытые интернет-архивы конференций, что снижает качество информации.

Расчитаны коэффициенты сходства Жаккара (в качестве меры сходства двух списков сериальных изданий) для разных хронологических периодов и различных ресурсов. Показано, что состав источников в тематической коллекции The Lens (Fields of Study) более динамичен во времени (30,72%), чем в коллекциях Dimensions и The Lens (Subjects), где коэффициенты Жаккара были, соответственно – 38,89% и 51,52%.

Таким образом, аналитические инструментариумы БД Dimensions и VOSViewer дают возможность выявлять журналы не только с высокой продуктивностью, но и с высокой средней цитируемостью статей.

Использование программы визуализации данных VOSViewer расширяет возможности инструментариумов рассмотренных баз данных, что позволяет использовать БД The Lens и Dimensions для получения перечней наиболее продуктивных сериальных изданий по тематическим коллекциям, которые могут дополнять друг друга.

## СПИСОК ЛИТЕРАТУРЫ

1. Лутай А.В., Любушко Е.Э. Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection // Национальная подписка. – 2022. – URL: [https://podpiska.rfbr.ru/storage/reports2021/2022\\_meta\\_quality.html](https://podpiska.rfbr.ru/storage/reports2021/2022_meta_quality.html) (дата обращения: 01.07.2025)
2. Delgado-Quirós L., Ortega J.L. Completeness degree of publication metadata in eight free-access scholarly databases // Quantitative Science Studies. – 2024. – Vol.5, № 1. – P. 31–49.
3. Basson I., Simard M. A., Ouangré Z. A., Sugimoto C. R., Larivière V. The effect of data sources on the measurement of open access: A comparison of Dimensions and the Web of Science // PLOS ONE. – 2022. – Vol. 17, № 3. – e0265545.
4. Гуреев В.Н., Мазов Н.А. Возрастание роли открытых библиографических данных в условиях ограничения доступа к коммерческим информационным системам // Управление наукой: теория и практика. – 2023. – Т. 5, № 2. – С. 49–76.
5. Мохначева Ю.В., Цветкова В.А. Возможные пути получения научной информации в новых условиях // Управление наукой: теория и практика. – 2023. – Т. 5, № 3. – С. 117–158.
6. Тургель И.Д., Чернова О.А. Открытые библиографические базы данных: в поисках альтернативы Scopus и Web of Science // Экономика науки. – 2024. – Т. 10, № 3. – С. 34–51.
7. Martín-Martín A., Thelwall M., Orduna-Malea E. et al. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and

OpenCitations' COCI: a multidisciplinary comparison of coverage via citations // Scientometrics. – 2021. – Vol. 126. – P. 871–906.

8. Singh V.K., Singh P., Karmakar M., Leta J., Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis // Scientometrics. – 2021. – Vol. 126, № 6. – P. 5113–5142.
9. Гуреев В.Н., Ильичёва И.Ю., Мазов Н.А. Профили авторов и организаций в информационных системах Dimensions и Lens: исследование возможностей // Научные и технические библиотеки. – 2023. – № 10. – С. 138–170.
10. Мохначева Ю. В. Тенденции в международном соавторстве российских ученых в 2019–2023 гг. по данным OpenAlex // Библиосфера. – 2025. – № 1. – С. 95–113.
11. Turgel I.D., Chernova O.A. Open Science Alternatives to Scopus and the Web of Science: A Case Study in Regional Resilience // Publications. – 2024. – Vol. 12, № 4. – 43.
12. van Eck N.J., Waltman L. Manual for VOSviewer, Version 1.6.19 / Centre for Science and Technology Studies, Leiden University.– Leiden, The Netherlands, 2023.
13. Федорец О.В., Солошенко Н.С. Исследование тематической и терминологической трансформации коллекций научных статей по междисциплинарным областям знания // Научно-техническая информация. Сер.2. – 2024. – № 12.– С.15-25; Fedorets O.V., Soloshenko N.S. Study of Thematic and Terminological Transformation of Collections of Scientific Articles in Interdisciplinary Fields of Knowledge // Automatic Documentation and Mathematical Linguistics. – 2024. – Vol. 58, № 6. – P. 420-430.
14. Wei-Lin Hsu, Che-Wei Tsai, An-Chou Yeh, Jien-Wei Yeh. Clarifying the four core effects of high-entropy materials // Nature Reviews Chemistry. – 2024. – Vol. 8 – P. 471-485.
15. Yeh J.W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes // Advanced. Engineering. Materials. – 2004. – Vol. 6 – P. 299–303.
16. Oses C., Toher C., Curtarolo S. High-entropy ceramics // Nature Reviews Materials. – 2020. – Vol. 5 – P. 295–309.

*Материал поступил в редакцию 27.08.25.*

## Сведения об авторах

**СОЛОШЕНКО Наталия Сергеевна** – кандидат педагогических наук, заведующий Отделом комплектования Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН), Москва  
e-mail: solns@viniti.ru

**ФЕДОРЕЦ Олег Владимирович** – кандидат технических наук, заведующий лабораторией средств автоматизации Отдела информационных технологий ВИНИТИ РАН  
e-mail: ovf@viniti.ru

## Эволюция или упадок? Критический анализ глобальных тенденций трансформации общедоступных библиотек

*Представляется критический анализ исследования В.К. Степанова, посвященного глобальным тенденциям развития общедоступных библиотек. Рассматриваются как достоинства источниковой базы публикации и широкое использование статистического метода, так и методологические ограничения работы, спорные интерпретации данных и теоретические упущения, связанные с узким пониманием трансформации библиотек. Автор настоящей статьи предлагает альтернативные гипотезы, объясняющие динамику библиотечных сетей не только цифровизацией, но и экономическими, политическими и социокультурными факторами. Особое внимание уделяется новым функциям библиотек в цифровую эпоху: их роли как центров достоверной информации, хранителей культурного наследия и площадок социального участия. В Заключении обосновывается необходимость переосмысления теоретических подходов к изучению библиотек с учетом их перехода от традиционных моделей к полифункциональным институтам.*

**Ключевые слова:** общедоступные библиотеки, цифровая трансформация, библиотечная статистика, социокультурные функции библиотек, государственная политика, библиотеки как центры знаний, критика библиотечных исследований

DOI: 10.36535/0548-0019-2025-11-6

### ВВЕДЕНИЕ

В последние десятилетия общедоступные библиотеки переживают сложный период трансформации, вызванный цифровой революцией, изменениями в моделях информационного потребления и трансформацией социальных институтов. Исследование В.К. Степанова\*, основанное на анализе библиотечной статистики 33-х стран, стало одной из немногих попыток систематизировать глобальные тенденции в этой сфере. Однако, несмотря на ценность собранных данных, работа вызывает ряд методологических и концептуальных вопросов, которые станут предметом нашего анализа в данной статье.

Автор сводит динамику библиотек к простой дихотомии «роста» и «сокращения», объясняя эти процессы преимущественно цифровизацией и уровнем экономического развития. Такой подход игнорирует комплексность современной трансформации библиотек, которые все чаще берут на себя функции центров образования, культурного наследия и социальной интеграции. Кроме того, исследование не учитывает влияние бюджетной политики, государственных программ и инициатив профессионального сообщества,

которые во многих случаях оказываются более значимыми факторами, чем технологические изменения.

Цель статьи – критически переосмыслить выводы В.К. Степанова, предложив более многомерный взгляд на эволюцию библиотек. Особое внимание уделяется альтернативным объяснительным моделям, а также новым ролям, которые библиотеки играют в условиях цифрового общества.

### ЗНАЧИМОСТЬ ИССЛЕДОВАНИЯ И ЕГО МЕТОДОЛОГИЧЕСКИЕ ДОСТОИНСТВА

Работа В.К. Степанова представляет собой важный вклад в изучение глобальных тенденций развития общедоступных библиотек, поскольку впервые на столь обширном материале (33 страны, представляющие разные континенты и уровни социально-экономического развития) систематизированы данные об их количественной динамике. Использование авторитетных источников — национальной статистики, данных IFLA, платформы Statista — придает исследованию высокую степень достоверности, что особенно ценно в условиях фрагментированности мировой библиотечной статистики. Автор не ограничивается простым сопоставлением цифр, а выявляет две ключевые, причем противоположные, тенденции: устойчивое сокращение числа библиотек в экономически развитых странах (Россия (отнесена автором к списку развитых стран), Германия, США, Великобритания)

\* Степанов В.К. Современные тренды эволюции общедоступных библиотек по данным мировой библиотечной статистики // Научно-техническая информация. Сер. 1. – 2025. – № 3. – С.11-21.

и их рост в государствах с догоняющей экономикой (Китай (рассматривается автором как развивающееся государство), Турция, Аргентина, Индонезия). Этот дуализм позволяет говорить о глобальном перераспределении роли библиотек в зависимости от уровня технологического развития общества и степени зрелости его информационной инфраструктуры.

Особую научную ценность имеет выделение стран-исключений — Австралии, Южной Кореи и Индонезии, — где, несмотря на высокий экономический уровень или стремительный рост, количество библиотек не только не уменьшается, но и увеличивается. Автор справедливо связывает эту аномалию с радикальным расширением их функций: трансформацией из традиционных центров книговыдачи в многофункциональные общественные пространства, сочетающие образовательные, коммуникационные и социальные задачи. Такой подход позволяет предположить, что будущее библиотек определяется не их формальным наличием в статистических отчетах, а способностью адаптироваться к новым запросам цифрового общества.

Методологическая строгость работы проявляется и в четком определении объекта исследования — учитывались только муниципальные библиотеки, финансируемые местными властями, что обеспечило сопоставимость данных. Кроме того, В.К. Степанов фиксирует проблему терминологической разрозненности (например, различия в понятиях «библиотечный пункт» в Нидерландах и Австралии), подчеркивая необходимость международной стандартизации учета. Таким образом, исследование не только предоставляет актуальную статистическую картину, но и задает рамки для дальнейших дискуссий о путях развития библиотечного дела в условиях цифровой трансформации.

Однако именно эта методологическая строгость одновременно становится и ограничением, поскольку фокус на количественных показателях оставляет за скобками качественные изменения в работе библиотек — их посещаемость, спектр услуг, взаимодействие с локальными сообществами. Тем не менее, работа В.К. Степанова формирует прочную основу для последующего анализа, демонстрируя, что судьба библиотек в разных странах определяется не универсальными законами, а сложным взаимодействием технологических, экономических и культурных факторов.

### **Методологические ограничения исследования**

Несмотря на несомненную ценность проведенного анализа, работа содержит ряд существенных методологических ограничений, которые влияют на интерпретацию результатов и выводов. Прежде всего, это проблема неоднородности данных. Хотя автор честно признает различия в критериях учета библиотек в разных странах (например, исключение сельских библиотек из статистики Индонезии или различия в определении «библиотечного пункта»), это существенно снижает надежность межстрановых сравнений. В случаях с Бразилией и Канадой прихо-

дится констатировать либо противоречивость официальных данных, либо их полное отсутствие, что вынуждает исследователя опираться на косвенные и зачастую неофициальные источники. Такая ситуация неизбежно вносит элемент субъективности в анализ и ставит под вопрос точность количественных сопоставлений.

Еще более серьезным ограничением является преобладание количественного анализа при недостаточном присутствии качественного. Исследование фиксирует динамику численности библиотек, но оставляет без внимания принципиально важные аспекты их трансформации: изменение посещаемости, появление новых услуг, модификацию функций в цифровую эпоху. Например, сокращение физической посещаемости библиотек в Европе может сопровождаться взрывным ростом виртуальных посещений и расширением онлайн-сервисов, однако эти компенсаторные механизмы остаются за рамками анализа. Такой односторонний подход создает искаженное представление о реальных процессах модернизации библиотечного дела, сводя сложную институциональную трансформацию к простой арифметике открытий и закрытий учреждений.

Существенную проблему представляет и избирательность выборки. Включение лишь 33-х стран из 193, при явном доминировании экономически развитых государств и быстрорастущих экономик, создает существенный перекокс в репрезентативности исследования. Особенно заметно это в отношении Африки (представленной лишь тремя странами) и значительной части Азии. Такой дисбаланс хотя и объясняется ограниченностью мировой библиотечной статистики, не позволяет в полной мере говорить о действительно глобальных тенденциях, а скорее фиксирует ситуацию в наиболее «удобных» для анализа странах с относительно развитой статистической системой. В результате формируется картина, где доминируют процессы, характерные для технологически развитых обществ, тогда как специфика библиотечного развития в наименее развитых странах остается в тени. Эти методологические лакуны существенно ограничивают универсальность выводов и требуют осторожности при их экстраполяции на глобальную библиотечную систему в целом.

### **СПОРНЫЕ ИНТЕРПРЕТАЦИИ В ИССЛЕДОВАНИИ**

Одним из наиболее уязвимых аспектов исследуемой работы являются предлагаемые автором интерпретации выявленных тенденций, которые зачастую носят излишне прямолинейный характер. Центральный тезис о том, что сокращение библиотек в развитых странах обусловлено исключительно переходом на цифровые ресурсы, выглядит некоторым упрощением реальной ситуации. Эта гипотеза не может объяснить, почему в таких технологически продвинутых странах как Южная Корея и Австралия наблюдается не сокращение, а устойчивый рост библиотечной сети. Более убедительной представляется альтернативная версия, связывающая динамику библиотечной сети с процессами бюджетной оптимизации и пере-

распределения ресурсов в условиях экономических кризисов и изменения приоритетов государственного финансирования. В пользу этого говорит и тот факт, что наиболее резкое сокращение библиотек в Европе пришлось на период после финансового кризиса 2008 г., когда многие правительства проводили политику жесткой экономии.

Аналогичные вопросы вызывает трактовка роста числа библиотек в развивающихся странах через призму «догоняющего развития». Предложенное объяснение выглядит слишком общим и не учитывает существенных различий между странами. Яркий пример - колоссальный разрыв между скромным ростом в Китае (+2,3%) и взрывным увеличением в Индонезии (+518%). Такая диспропорция явно указывает, что за этими процессами стоят не абстрактные законы модернизации, а конкретные национальные стратегии и программы библиотечного развития. В случае Индонезии решающую роль сыграла целенаправленная государственная политика по созданию сети сельских библиотек как инструмента социального развития, тогда как в Китае акцент делался на качественное преобразование уже существующей сети.

Особого внимания заслуживает анализ российской ситуации, где автор ограничивается ссылкой на цифровизацию как главную причину сокращения сети библиотек на 27,6%. Такой подход полностью игнорирует важные региональные особенности: массовое закрытие сельских библиотек в первую очередь связано с процессами депопуляции сельских территорий и оптимизацией бюджетных расходов, а не с переходом на цифровые форматы. Не менее значимым фактором являются проводимые реформы централизации, ведущие к укрупнению библиотечных сетей и сокращению числа юридически самостоятельных учреждений при сохранении фактического объема услуг. Эти нюансы принципиально важны для понимания реальной динамики библиотечной системы, но остаются за рамками исследования. В результате создается несколько упрощенная и во многом искаженная картина происходящих изменений.

В исследовании В.К. Степанова прослеживается существенный концептуальный пробел, связанный с узким пониманием трансформации библиотек исключительно через призму их информационной функции. Такой подход не учитывает глубинных изменений, превращающих современные библиотеки в сложные социокультурные институты нового типа. Автор упускает из виду, что в условиях цифровой эпохи библиотеки берут на себя принципиально новые роли, выходящие далеко за рамки традиционного книгохранения. Они становятся центрами верификации информации в мире фейков и цифрового шума, хранителями цивилизационного кода через оцифровку культурного наследия, площадками формирования идеологических ориентиров и гражданской идентичности. Особенно важна их трансформирующаяся роль как пространств социального участия, где через образовательные программы, дискуссионные клубы и локальные инициативы формируется новая культура общественной вовлеченности.

Эта качественная метаморфоза библиотечных институтов приводит к появлению принципиально новых организационных форм, которые уже трудно назвать «библиотеками» в классическом понимании. В разных культурных контекстах они эволюционируют в уникальные гибридные модели: в скандинавских странах – в культурно-просветительские хабы, объединяющие функции музея, лектория и коворкинга; в Южной Корее – в центры цифровой грамотности, где обучение работе с информацией превалирует над традиционной книговыдачей; в странах Латинской Америки – в общественные резиденции, аккумулирующие самые разные локальные инициативы. Особенно показательно, что пандемия COVID-19, полностью выпавшая из поля зрения исследования, стала мощным катализатором этих процессов, превратив гибридные форматы библиотечного обслуживания из временной меры в новую институциональную норму.

Игнорирование этой комплексной трансформации существенно обедняет анализ, сводя сложный процесс институциональной эволюции к упрощенной количественной дихотомии «роста и сокращения». Между тем, современная библиотечная реальность требует принципиально новых теоретических подходов, способных осмыслить библиотеки как полифункциональные платформы социокультурного развития. Такой концептуальный сдвиг предполагает переосмысление не только функций, но и самой сущности библиотек, которые во многих случаях уже перестают быть собственно библиотеками в традиционном понимании, превращаясь в уникальные синтетические институты, сочетающие черты культурных центров, образовательных площадок и пространств общественной самоорганизации.

В рамках рассматриваемого исследования стоило бы рассмотреть альтернативные объяснительные модели, которые позволяют преодолеть ограниченность авторской трактовки. Прежде всего, тезис о «сокращении библиотек» требует принципиального переосмысления – современные реалии показывают, что мы имеем дело не с исчезновением института библиотек как такового, а с его качественной трансформацией. В развитых странах уменьшение количества отдельных библиотечных помещений компенсируется радикальным расширением функционала оставшихся учреждений, которые превращаются в многослойные центры образования и социализации. Эти новые пространства органично сочетают традиционные книжные фонды с лекториями, коворкингами, цифровыми лабораториями и общественными площадками, что принципиально меняет саму природу библиотечного обслуживания. При этом экономические факторы зачастую играют более значимую роль, чем технологические изменения – анализ бюджетных ассигнований в США и странах ЕС демонстрирует, что сокращение библиотечной сети часто становится следствием финансовой оптимизации, а не объективного снижения потребности в библиотечных услугах.

Парадоксальным образом в развивающихся странах рост числа библиотек также обусловлен в первую очередь политическими и экономическими фактора-

ми, а не собственно информационными потребностями. Активное библиотечное строительство в Индонезии, Турции или отдельных регионах Латинской Америки связано прежде всего с государственными программами развития инфраструктуры, где библиотекам отводится роль символов модернизации и центров локального просвещения. Особого внимания заслуживает позиция профессионального библиотечного сообщества, которое автор необоснованно обвиняет в пассивности. На практике именно библиотекари становятся инициаторами инновационных трансформаций – от внедрения цифровых платформ до создания принципиально новых форматов работы с местными сообществами. Их стратегии адаптации, включающие перепрофилирование пространств, разработку новых услуг и активное лоббирование интересов отрасли, заслуживают отдельного глубокого исследования как важнейший фактор эволюции библиотек в XXI в. Эти альтернативные перспективы анализа позволяют выйти за рамки противопоставления «цифрового» и «традиционного», раскрывая сложную диалектику трансформации библиотек как социального института.

## **ЗАКЛЮЧЕНИЕ**

Проведенный анализ показывает, что исследование В.К. Степанова, хотя и предоставляет ценный статистический материал, имеет некоторые недостатки необоснованного редуционизма. Сведения трансформации библиотек исключительно к количественным изменениям и технологическим факторам

не позволяет в полной мере понять их современную эволюцию.

На деле библиотеки не исчезают, а перерождаются, принимая на себя новые функции – от центров цифровой грамотности до площадок гражданской активности. Их сокращение в развитых странах часто связано не с отсутствием читательского спроса, а с бюджетными ограничениями и реструктуризацией, тогда как рост в развивающихся государствах обусловлен скорее политическими решениями, чем спросом на традиционные библиотечные услуги.

Перспективы дальнейших исследований трансформации общедоступных библиотек видятся в разработке более гибких теоретических моделей, учитывающих полифункциональность современных библиотек, их адаптационные стратегии и роль в формировании информационной и культурной устойчивости общества. Библиотеки будущего – это уже не просто хранилища книг, а динамичные социокультурные хабы, и их изучение требует междисциплинарного подхода, сочетающего методы библиотековедения, социологии и цифровой антропологии.

*Материал поступил в редакцию 04.08.25.*

## **Сведения об авторе**

**СОКОЛОВ Сергей Валерьевич** – заведующий научно-исследовательским отделом библиотековедения Института научной информации по общественным наукам РАН, Москва.  
e-mail: beholder73@gmail.com