

## ОТКРЫТЫЙ ДОСТУП / OPEN ACCESS

Оригинальная статья / Original paper

<https://doi.org/10.24069/SEP-251046>

## Ресурсы открытого доступа: эффективность метаданных и сервисного функционала

Н. С. Редькина *Государственная публичная научно-техническая библиотека СО РАН,  
г. Новосибирск, Российская Федерация*✉ [redkina@spsl.nsc.ru](mailto:redkina@spsl.nsc.ru)

**Резюме.** Развитие идей и принципов открытой науки способствовало распространению ресурсов открытого доступа (РОД), появлению широкого спектра информационно-поисковых и библиометрических возможностей для анализа различных документов, журналов, авторов, организаций, направлений исследований. Это стало возможным благодаря повышению эффективности метаданных за счет установления связей между ресурсами, внедрения в РОД собственных классификационных систем и обогащения данных с помощью технологий искусственного интеллекта. Метаданные играют важную роль в повышении эффективности РОД, так как являются средствами управления документами, обеспечивая их идентификацию, систематизацию, видимость, поиск, обнаружение и контроль доступа в информационно-поисковых системах. Однако не выработаны единые подходы к решению ключевых задач, связанных с функционированием ресурсов: 1) отражению метаданных в РОД; 2) разработке классификационных схем и сервисов для обработки результатов поиска; 3) представлению поисковых полей. Это усугубляет проблемы, часто возникающие при работе с информационными ресурсами, связанные с качеством метаданных (неполнота, неточность, ошибки, несогласованность, субъективность предметизации и систематизации) и их интероперабельностью при автоматическом обмене, обработке и использовании. Цель данного исследования — проанализировать эффективность представления метаданных в крупнейших научных РОД и функционал этих ресурсов для решения задач поиска и обработки документов. Для достижения поставленной цели на первом этапе был проведен контент-анализ 10 обзорных публикаций за 2023–2025 гг., в которых сравнивались крупнейшие научные РОД (Crossref, Dimensions, OpenAlex, Scilit, Semantic Scholar, The Lens и др.) с международными наукометрическими базами данных Scopus и Web of Science Core Collection на предмет полноты и качества метаданных. Результаты контент-анализа подтвердили возможности использования РОД в качестве библиографических источников поиска и анализа научной информации. Для верификации полученных данных на втором этапе были выявлены и изучены метаданные в четырех выбранных крупнейших научных РОД (OpenAlex, Dimensions, The Lens и Semantic Scholar), оценен их сервисный функционал для проведения библиометрических исследований и визуализации полученных данных. Далее были определены собственные либо используемые системы классификации документов в РОД и рассмотрены особенности выгрузки больших объемов метаданных. Сравнительный анализ OpenAlex, Dimensions, The Lens и Semantic Scholar показал различия по полноте полей, аналитическим сервисам и ограничениям экспорта, а также типовые уязвимости качества метаданных. Сделан вывод о перспективах развития этих ресурсов для поиска научной информации и проведения исследований с учетом постоянно развивающегося функционала и использования надежных источников, таких как Crossref, ORCID, ROR и др., для пополнения РОД. Наличие международных идентификаторов, устранение неоднозначности в описании, решение проблем индексирования документов при присвоении дополнительных полей метаданных позволят использовать РОД как эффективные информационные ресурсы. Это будет способствовать формированию более прозрачного исследовательского ландшафта, многопараметрическому анализу

развития научных направлений и сотрудничества, углубленному изучению тематик исследований, повышению видимости изданий в мировом информационном пространстве, а также цитируемости авторов, научных организаций и журналов благодаря отражению в большем количестве связанных ресурсов, дополнительным метаданным для публикации, а также сервисным возможностям и функционалу ресурсов для дальнейшей обработки и представления данных с помощью средств визуализации, инструментов построения карт знаний и других технологий.

**Ключевые слова:** открытая наука, ресурсы открытого доступа, научные журналы, научные публикации, метаданные, информационный поиск, идентификатор ученого, ORCID, идентификатор организации, ROR, идентификатор цифрового объекта, DOI, альтметрика, библиометрические показатели, визуализация данных

**Для цитирования:** Редькина Н. С. Ресурсы открытого доступа: эффективность метаданных и сервисного функционала. *Научный редактор и издатель*. 2025;10(2):289-306. <https://doi.org/10.24069/SEP-251046>

---

## Open access resources: the efficiency of metadata and services

N. S. Redkina 

*State Public Scientific Technological Library, Siberian Branch of the Russian Academy of Sciences,  
Novosibirsk, Russian Federation*

✉ [redkina@spsl.nsc.ru](mailto:redkina@spsl.nsc.ru)

---

**Abstract.** The development of open science ideas and principles has facilitated the spread of open access resources (OARs), creating a wide range of information retrieval and bibliometric capabilities for analyzing various documents, journals, authors, organizations, research areas, thanks to the increased effectiveness of metadata, the links between resources, the implementation of proprietary classification systems in OARs, and data enrichment using artificial intelligence technologies. Metadata plays a crucial role in enhancing the effectiveness of OARs, as it serves as a means of document management, ensuring their identification, systematization, visibility, search, discovery, and access control in information retrieval systems. However, unified approaches are lacking regarding 1) the representation of metadata in OARs, 2) the development of classification schemes and services for processing search results, and 3) the presentation of search fields. This exacerbates the problems often encountered when working with information resources related to metadata quality (incompleteness, inaccuracy, errors, inconsistencies, subjectivity in subjectification and systematization) and their interoperability during automatic exchange, processing, and use. The study aims to analyze the effectiveness of metadata representation in major scientific OARs and their functionality for document search and processing. To achieve this goal, the first stage involved a content analysis of 10 reviews from 2023–2025. These reviews compared major scientific OARs (Crossref, Dimensions, OpenAlex, Scilit, Semantic Scholar, The Lens, etc.) with international scientometric databases Scopus and Web of Science Core Collection in terms of metadata completeness and quality, confirming the potential of OARs as bibliographic sources for scientific information retrieval. To validate these findings, the second stage involved an analysis of metadata in four selected major scientific OARs (OpenAlex, Dimensions, The Lens, and Semantic Scholar), assessing their service functionality for conducting bibliometric research and visualizing the resulting data. Furthermore, the proprietary or adopted document classification systems within the OARs were examined, along with the specifics of exporting large volumes of metadata. A comparative analysis of OpenAlex, Dimensions, The Lens, and Semantic Scholar revealed differences in field completeness, analytical services, export limitations, as well as typical metadata quality vulnerabilities. Nevertheless, the prospects for the development of these resources for scientific information retrieval and research are acknowledged, considering their continuously evolving functionality and the use of reliable sources for data enrichment, such as Crossref, ORCID, ROR, etc. The presence of international identifiers, the resolution of ambiguities in description, and addressing issues in document indexing when assigning additional metadata fields in OARs will enable their use as effective information resources. This will contribute to forming a more transparent research landscape, enable multiparameter analysis of the development of scientific fields and collaboration, facilitate multifaceted study of research topics, and

enhance the visibility of publications in the global information space. It will also increase the citation potential of authors, scientific organizations, and journals through inclusion in a greater number of interconnected resources, provision of additional metadata for publication, and the availability of service capabilities and functionalities for further processing and presentation using visualization tools, knowledge mapping instruments, and others.

**Keywords:** open science, open access resources, scientific journals, scientific publications, metadata, information retrieval, researcher identifier, ORCID, organization identifier, ROR, digital object identifier, DOI, altmetrics, bibliometric indicators, data visualization

**For citation:** Redkina N.S. Open access resources: the efficiency of metadata and services. *Science Editor and Publisher*. 2025;10(2):289-306. <https://doi.org/10.24069/SEP-251046>

## ВВЕДЕНИЕ

Динамичное развитие инфраструктуры открытой науки способствует появлению большого количества научных ресурсов открытого доступа (РОД), в которых фиксируются разнообразные результаты исследований. Научные РОД, такие как Dimensions, OpenAlex, The Lens, архивы препринтов (arXiv, bioRxiv, [PREPRINTS.RU](https://preprints.ru)), репозитории данных (FAIRsharing, Zenodo), научные поисковые системы (Google Scholar, Semantic Scholar) и др., привлекают все большее внимание пользователей и библиотечных специалистов благодаря доступности, расширенным параметрам для поиска и анализа, отражению разных видов документов и наборов данных, включению в эти РОД метрик и различных дополнительных сервисов. Их совершенствованию способствуют инициативы по открытию исследовательской информации, в том числе с помощью разнообразных метаданных.

Опубликованная в 2024 г. Барселонская декларация об открытой научной информации<sup>1</sup> содержит требование к издательским сервисам и платформам обеспечивать открытый доступ через открытые инфраструктуры с использованием стандартных протоколов и идентификаторов метаданных научных статей и других результатов. В Декларации указано, что такие метаданные (например, названия, аннотации, наименования журналов, информация об авторах и их аффилиациях, финансирующих организациях и т.д.) в известных международных наукометрических базах данных (МНБД) Web of Science (WoS) и Scopus имеют серьезные ограничения в использовании, так как доступны только для организаций, оплачивающих подписку. Это, в свою очередь, снижает прозрачность и воспроизводимость показателей и аналитики, основанных на этих базах данных (БД) (например, статистики публикаций и цитирования, импакт-факторов журналов, рейтингов

университетов и т.д.). В Рекомендации по открытой науке ЮНЕСКО также подчеркивается важность открытых библиометрических и наукометрических систем для оценки и анализа научных областей<sup>2</sup>.

Открытость исследовательской информации, в частности, метаданных научных статей поддерживают также Инициатива за открытые цитирования<sup>3</sup> и Инициатива за открытые рефераты<sup>4</sup>, а также проект «Метаданные 20/20»<sup>5</sup>. В 2025 г. создана Объединенная целевая группа по вопросам обеспечения открытости метаданных публикаций<sup>6</sup>, направляющая свои усилия на достижение большей их прозрачности. В этом же году появилась инициатива, призванная повысить полноту, согласованность и совместимость метаданных в экосистеме научной информации, — COMET (Collaborative Metadata Enrichment Taskforce — Рабочая группа по совместному обогащению метаданных, в дальнейшем переименованная в «Совместные метаданные»)<sup>7</sup>. COMET предлагает модель курирования, позволяющую широкому кругу заинтересованных сторон совместно обогащать метаданные, поскольку их пользователи, такие как исследователи, учреждения, спонсоры и агрегаторы, часто обладают информацией, которая могла бы улучшить эти записи.

<sup>2</sup> UNESCO Recommendation Concerning Open Educational Resources. 40 C/32 Annex. Paris: UNESCO, 2019. URL: <https://opening-up.education/wp-content/uploads/2019/12/RECOMMENDATION-CONCERNING-OPEN-EDUCATIONAL-RESOURCES.pdf> (дата обращения: 03.12.2025).

<sup>3</sup> Initiative for Open Citations (I4OC). 2022. URL: <https://i4oc.org/> (дата обращения: 03.12.2025).

<sup>4</sup> Initiative for Open Abstracts, 2022. URL: <https://i4oa.org> (дата обращения: 03.12.2025).

<sup>5</sup> Metadata 20/20. URL: <https://metadata2020.org/> (дата обращения: 03.12.2025).

<sup>6</sup> Barcelona Declaration and OA2020 launch Joint Task Force on Negotiating Openness of Publication Metadata. 2025. URL: [https://barcelona-declaration.org/news/20251002\\_bd\\_oa2020\\_joint\\_taskforce/](https://barcelona-declaration.org/news/20251002_bd_oa2020_joint_taskforce/) (дата обращения: 03.12.2025).

<sup>7</sup> COMET. URL: <https://www.cometadadata.org> (дата обращения: 03.12.2025).

<sup>1</sup> Barcelona Declaration on Open Research Information. URL: <https://barcelona-declaration.org> (дата обращения: 03.12.2025).

Основная идея перечисленных инициатив заключается в том, что создание более полных, взаимосвязанных и многократно используемых открытых метаданных для всех результатов исследований способствует развитию науки. Метаданные являются средством управления документами, обеспечивая их идентификацию, систематизацию, видимость, поиск, обнаружение и контроль доступа в информационно-поисковых системах. РОД играют важную роль в обеспечении открытости и доступности научных публикаций, что актуализирует значение качественных метаданных. Они обеспечивают интероперабельность (способность разных информационных систем обмениваться информацией), способствуют повторному использованию результатов исследований, повышают видимость публикаций и источников, улучшают последующее цитирование в других документах, а также соцсетях, блогах, средствах массовой информации. Повышение качества метаданных в средах больших данных позволяет оптимизировать время и стоимость обработки данных [1]. Метаданные с течением времени меняются, появляются новые поисковые поля в РОД, такие, например, как идентификаторы авторов и организаций, статус открытого доступа, финансирующие организации, собственные классификаторы дисциплин и отраслей знаний, библиометрические и альтметрические показатели и др. [2–4].

Несмотря на динамичное развитие РОД, которые обеспечивают широкий библиографический и полнотекстовый охват научной литературы и расширяют состав метаданных, сохраняется вопрос о достоверности этих метаданных, ограничениях при поиске и идентификации документов [5]. Исследования демонстрируют, что БД, работающие на основе внешних источников (Dimensions, OpenAlex, Scilit и The Lens), могут формировать более полные и качественные метаданные, чем академические поисковые системы (Google Scholar, Semantic Scholar), извлекающие информацию из интернета и не позволяющие получить надежные описательные данные о документах [6]. Основными проблемами РОД остаются качество и структура метаданных: их полнота, точность (отсутствие опечаток, единообразие в написании личных имен и топонимов, использование стандартных сокращений), соответствие стандартам, доступность и другие аспекты. Учитывая, что корректные метаданные во многом определяют дизайн, масштаб и качество исследования, важно определить сильные и слабые стороны их представления в РОД. Целью настоящей статьи стал анализ эффективности представления метаданных и функционала

сервисов в крупнейших научных РОД с точки зрения поисковых возможностей, наличия собственных классификаций, доступности, разнообразия и полноты метаданных.

## МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

### Материалы

В результате анализа РОД, представленных на сайте «Библиотека для открытой науки» (<https://lib-os.ru/issledovatelyam/resursy/>), выявлены крупнейшие по охвату документов и универсальные по содержанию международные информационно-поисковые системы (данные на ноябрь 2025 г.):

- OpenAlex (<https://openalex.org/>) — более 271 млн документов;
- Dimensions (<https://app.dimensions.ai/>) — более 160 млн публикаций из свыше 70 000 журналов; 62 млн наборов данных из более чем 100 ведущих мировых репозиторий; более 176 млн полнотекстовых патентных записей и др.;
- Semantic Scholar (<https://www.semanticscholar.org>) — более 230 млн документов;
- Google Scholar (<https://scholar.google.com/>) — данные о количестве документов не раскрываются;
- BASE, Bielefeld Academic Search Engine (<https://www.base-search.net/>) — более 460 млн документов, из них более 60% в открытом доступе;
- The Lens (<https://www.lens.org/>) — свыше 270 млн документов;
- CORE (<https://core.ac.uk/>) — более 405 млн документов.

Исходя из цели исследования определены РОД, предоставляющие возможности не только для поиска литературы, отслеживания ссылок, но и для выявления закономерностей цитирования, оценки исследования и извлечения больших выборок метаданных с помощью легальных API (Application Programming Interface — интерфейс прикладного программирования) или путем экспорта значительных объемов результатов поиска в виде файлов специальных форматов (например, BibTeX, RIS) для последующего импорта в библиографические менеджеры (Zotero, Mendeley, EndNote, RefWork и др.) с целью анализа и визуализации. В связи с этим, несмотря на популярность среди пользователей и широкий охват различных видов документов, Google Scholar был исключен из анализа (по причине отсутствия открытого API).

Эмпирической базой для достижения поставленной цели послужили две группы источников: 1) Научная электронная библиотека (eLIBRARY.RU) и OpenAlex — для поиска и анализа обзорных публикаций, в которых сравниваются научные РОД с МНБД

Scopus и WoS; 2) выбранные научные РОД (OpenAlex, Dimensions, The Lens и Semantic Scholar) — для выявления особенностей и возможностей использования метаданных в целях дальнейшего анализа и визуализации.

### Методы и инструменты

Из 34 крупнейших мировых РОД для анализа было отобрано 4, соответствующих цели исследования. Выбранные ресурсы были изучены по следующим аспектам: предлагаемым поисковым полям, отражаемым метаданным, используемым технологиям и объемам выгрузки метаданных (через API или экспорт данных в различных форматах для последующей обработки с помощью встроенных или внешних аналитических инструментов). В работе применялись современные цифровые инструменты выгрузки, визуализации и анализа данных: 1) табличный редактор MS Excel для визуализации динамики пополнения OpenAlex; 2) БД Wikidata в целях анализа классификационных понятий, применяемых в OpenAlex, с оценкой их релевантности; 3) API для массового поиска в Semantic Scholar и The Lens и выгрузки документов из этих РОД; 4) VOSviewer для визуализации библиографических связей в цитированиях российских патентов (по данным The Lens).

### Процедуры анализа

Анализ осуществлялся в несколько этапов. На первом этапе проведен тематический поиск обзорных публикаций по теме «Сравнение метаданных различных ресурсов». Поисковое предписание было составлено с использованием фильтров (тип — статья, год — диапазон лет — 2021–2025) и включало в поисковое предписание названия крупнейших ресурсов (CrossRef, OpenAlex, Scopus, WoS, Google Scholar, Dimensions, РИНЦ, Semantic Scholar, The Lens). Выявлено более 30 статей, из которых отобрано 10 обобщающих работ 2023–2025 гг., наиболее релевантных запросу, для дальнейшего контент-анализа и изучения специалистами результатов исследований, сравнивающих РОД с известными МНБД и между собой. На втором этапе выявлены и проанализированы используемые метаданные и возможности расширенного поиска в анализируемых РОД (фильтры, булевы операторы, ограничения). На третьем этапе изучены собственные либо используемые системы систематизации/классификации документов в РОД. И, наконец, на заключительном этапе рассмотрены особенности выгрузки больших объемов метаданных для последующего анализа средствами табличных редакторов и визуализации в выбранных РОД.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### Контент-анализ 10 публикаций, отражающих результаты сравнительного анализа полноты и качества метаданных в крупнейших информационных ресурсах

В последние годы проводились сравнительные исследования полноты и качества метаданных, представленных в РОД и МНБД Scopus и WoS. В табл. 1 обобщены основные выводы из работ 2023–2025 гг., посвященных сравнительному анализу метаданных крупнейших информационных ресурсов, который позволил выявить сильные и слабые стороны ключевых БД, уточнить набор критериев и показателей оценки эффективности метаданных в соответствии с целью исследования, понять текущие тенденции развития крупнейших РОД.

Контент-анализ выбранных 10 обзорных публикаций (2023–2025 гг.), посвященных сравнению РОД (OpenAlex, Dimensions, The Lens и др.) с МНБД Scopus и WoS, позволил сделать ряд выводов: основные исследуемые параметры РОД — охват публикаций, качество метаданных, возможности ссылочного анализа, поддержка открытого доступа и функциональность. Основное внимание уделяется оценке их пригодности для научно-исследовательской работы, библиометрического анализа и институционального учета публикационной активности.

В работах подчеркиваются преимущества OpenAlex по сравнению с другими ресурсами: более широкое лингвистическое покрытие, сопоставимый охват ссылок, больший охват ORCID, открытость и доступность. К недостаткам этого ресурса относится более низкое качество метаданных и точность классификации метаданных; меньшее количество аннотаций и большее число ошибок в цитированиях, что снижает его надежность для наукометрических исследований и оценочной аналитики.

Несмотря на сохранение преимуществ WoS и Scopus в качестве, полноте и стандартизации метаданных, что критически важно для точных оценочных библиометрических расчетов, крупнейшие научные РОД уже рассматриваются как их замена в связи с высокой стоимостью подписки на эти коммерческие МНБД и совершенствованием открытых ресурсов [5; 16–18]. Показательно решение Университета Сорбонны (Франция) прекратить подписку на БД WoS и библиометрические инструменты Clarivate с 2024 г. и развивать взаимодействие с научной поисковой системой OpenAlex<sup>8</sup>.

<sup>8</sup> Sorbonne University unsubscribes from the Web of Science. Sorbonne University. 2023. URL: <https://www.sorbonne-universite.fr/en/news/sorbonne-university-unsubscribes-web-science> (дата обращения: 03.12.2025).

**Таблица 1.** Сравнение полноты и качества метаданных в крупнейших информационных ресурсах (по публикациям 2023–2025 гг.)**Table 1.** Comparison of the completeness and quality of metadata in the largest information resources (based on publications from 2023–2025)

Сравниваемые БД	Основные выводы по результатам исследований	Авторы [источник]
Dimensions, Google Scholar, Microsoft Academic, OpenAlex, Scilit, Semantic Scholar, The Lens	Проанализировано семь элементов метаданных (аннотация, доступ, библиографическая информация, тип документа, дата публикации, язык и идентификаторы) с точки зрения полноты этих полей и согласованности между БД. Вывод: академические поисковые системы (Google Scholar, Microsoft Academic и Semantic Scholar) собирают меньше информации и имеют низкую степень полноты, а БД Dimensions, OpenAlex, Scilit и The Lens имеют более высокое качество и полноту метаданных.	Л. Делгадо-Кирос и Дж.Л. Ортега (L. Delgado-Quirós, J.L. Ortega) [6]
OpenAlex, WoS	OpenAlex демонстрирует более сбалансированное лингвистическое покрытие, чем WoS (больше неанглоязычных статей). Ресурс хорошо конкурирует по количеству ссылок на публикации, но уступает в точности классификации и корректности управления метаданными.	Л. Сеспедес с соавт. (L. Céspedes et al.) [7]
OpenAlex, WoS, Scopus	OpenAlex показывает средние значения количества ссылок на статью, сопоставимые с WoS и Scopus (существенных отклонений не обнаружено); охватывает больше идентификаторов ORCID и данных о статусе открытого доступа, но содержит меньше аннотаций и больше неточностей в ссылках.	Дж. Кулберт с соавт. (J. Culbert et al.) [8]
AMiner, Wisdom.ai, The Lens, Dimensions, OpenAlex	Проведен сравнительный анализ выбранных наукометрических БД по полноте (количество источников различных типов) и временному охвату данных; способности характеризовать научное сообщество, выявлять наиболее авторитетные страны и университеты в данной области, а также существующие формы сотрудничества, центры научных знаний (страны и университеты), ведущие журналы по данной тематике. Вывод: РОД отражают тенденции развития науки, дублируя, дополняя и расширяя данные за счет различных источников, но не гарантируют высокого качества материалов и демонстрируют относительно низкий уровень метаданных. Преждевременно полностью отказываться от Scopus и WoS.	И. Д. Тургель и О.А. Чернова (Irina D. Turgel, Olga A. Chernova) [9]
WoS, Scopus, РИНЦ, Dimensions, The Lens	Dimensions и The Lens могут рассматриваться как полноценная замена коммерческим БД WoS и Scopus в качестве источников библиографической информации при формировании институциональных БД и репозиториях. Преимущества Dimensions и The Lens — в широком охвате научного контента, в том числе российских публикаций, оригинальных версий российских журналов и материалов конференций. Качество библиометрических показателей не сравнивалось.	Н. А. Мазов, В. Н. Гуреев [10; 11]
OpenAlex, Dimensions	OpenAlex ограничен в заполнении полей метаданных о финансирующих организациях, так как зависит от Crossref. Dimensions анализирует раздел «Благодарности», используя обработку естественного языка, чтобы определить спонсоров и обогатить метаданные из Crossref. У этого ресурса есть соглашения с издателями на получение дополнительной информации о финансировании.	Е. Шапес (E. Schares) [12]

Окончание табл. 1

Сравниваемые БД	Основные выводы по результатам исследований	Авторы [источник]
WoS, Scopus, Dimensions, OpenAlex	Dimensions и OpenAlex обладают новыми функциями, обеспечивающими лучший доступ и читаемость, и имеют более широкий тематический охват. Однако существуют проблемы с полнотой и надежностью данных. Эти БД подходят для поиска информации, но не для оценочной наукометрии.	П. Сингх с соавт. (P. Singh et al.) [13]
CrossRef, OpenCitations, OpenAlex, Scopus, WoS	Качество метаданных OpenAlex (аннотации, авторы, учреждения, ссылки и финансирование) ниже, чем у Scopus и WoS, особенно для публикаций, индексируемых исключительно в OpenAlex.	А. Велес-Эстевес (A. Velez-Estevez et al.) [14]
Scopus, WoS, Google Scholar, Dimensions, OpenAlex, Academia.edu, ResearchGate и др.	Google Scholar обладает широким библиографическим и полнотекстовым охватом. Библиографический охват Google Scholar и OpenAlex превосходит охват нескольких БД Манхэттенского университета. Метаданные, предоставляемые американскими социальными сетями и пиратскими сайтами, не соответствуют стандартам большинства библиотечных БД.	У. Х. Волтерс (W.H. Walters) [5]
Google Scholar, ResearchGate, Semantic Scholar, The Lens, Scopus, WoS, Scite, PubMed и JSTOR	Результаты анализа 59 БД помогают ученым, проводящим систематические обзоры, метаанализы и библиометрические исследования, выбрать наиболее подходящие БД для поиска цитирований. Google Scholar, ResearchGate, Semantic Scholar и The Lens являются основными системами для поиска данных и прямых цитирований (FWCI*), причем The Lens обеспечивает лучшие возможности загрузки.	М. Гузенбауэр (M. Gusenbauer) [15]

\* Показатель FWCI (Field-Weighted Citation Impact — «взвешенное по области знания цитирование») — это отношение суммарного количества цитирований, фактически полученных работами всех представителей категории, к общему количеству цитирований, ожидаемому исходя из среднего показателя для соответствующей отрасли знания.

Будущее РОД зависит от улучшения качества метаданных, активной поддержки издателями открытых инициатив, широкого внедрения идентификаторов (ORCID, ROR), повышения точности и надежности информации. Общий вывод согласуется с итоговим заключением А.В. Лутая и Е.Э. Любушко: если издатели продолжают поддерживать инициативы по переводу метаданных статей в открытый доступ, а организации станут больше использовать ROR и ORCID, качество и востребованность метаданных в РОД могут приблизиться к уровню известных коммерческих ресурсов<sup>9</sup>.

## Метаданные в РОД

### OpenAlex

Ресурс OpenAlex (<https://openalex.org/>) активно развивается и завоевал популярность во всем мире как востребованный открытый источник информа-

ции. По состоянию на сентябрь 2025 г. в OpenAlex представлено более 271 млн документов. Ежегодный прирост, начиная с 2010 г., составляет свыше 9 млн единиц (статьи, главы из книг, препринты и др.) (рис. 1). Крупнейшими источниками пополнения OpenAlex являются уже зарекомендовавшие себя ресурсы: Crossref, arXiv, PubMed, ORCID, ROR, DOAJ (Directory of Open Access Journal), Unpaywall, Internet Archive и др.

Все больше исследователей использует OpenAlex в качестве источника данных [19–22]. Этот ресурс обладает двумя основными преимуществами: легким доступом к информации и широким ее охватом, что делает его полезным инструментом для количественных научных исследований [23; 24]. В некоторых областях, например «сестринское дело» и др., OpenAlex индексирует почти все журналы, представленные в Scopus и WoS, а также дополнительные источники из стран за пределами западного мира [25; 26]. В исследованиях делается вывод, что OpenAlex можно рассматривать как расширенную версию Scopus и надежную альтернативу для определенных видов библиометрического анализа, особенно на уровне стран [18]. Так, рейтинг

<sup>9</sup> Лутай А.В., Любушко Е.Э. Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection. Отчет. Российский фонд фундаментальных исследований. URL: [https://podpiska.rcsi.science/storage/reports2021/2022\\_meta\\_quality.html](https://podpiska.rcsi.science/storage/reports2021/2022_meta_quality.html) (дата обращения: 03.10.2025).

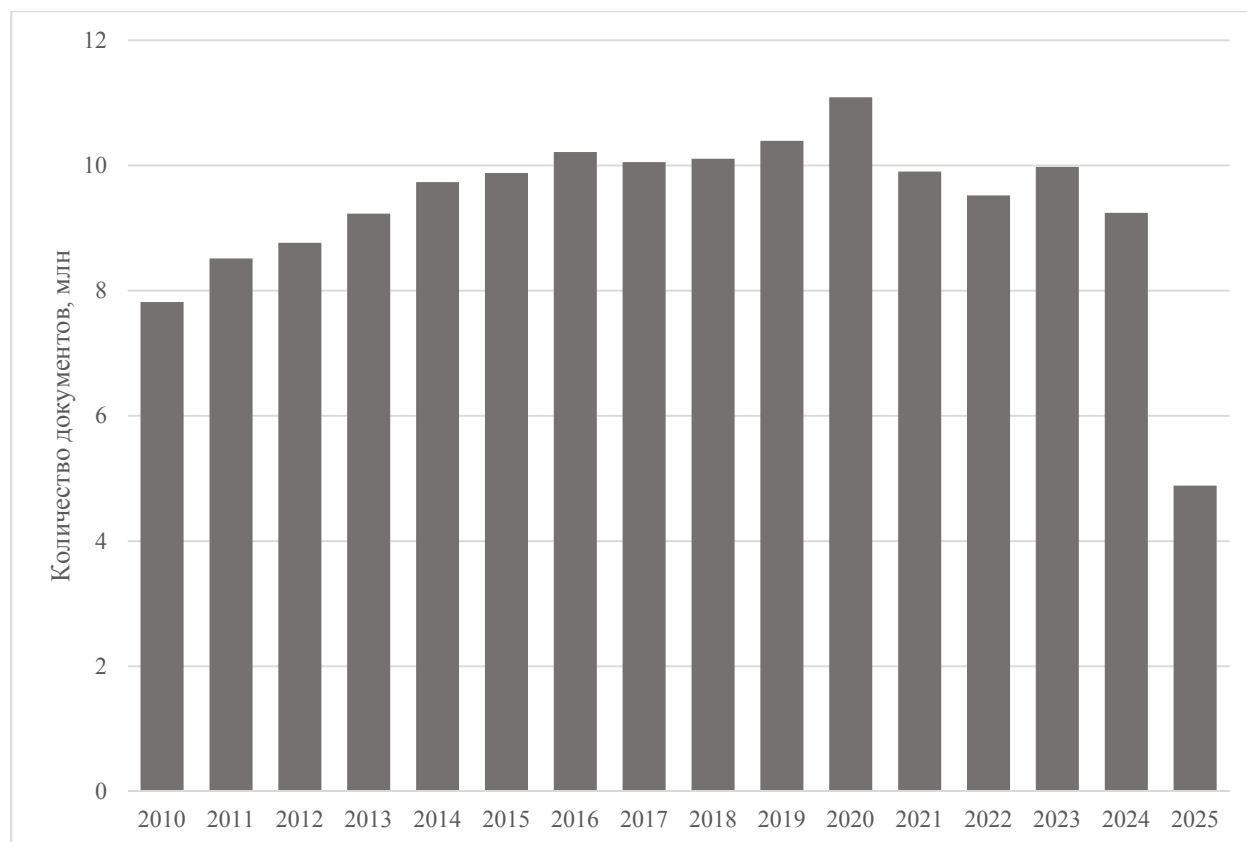


Рис. 1. Динамика пополнения OpenAlex (2010 — сентябрь 2025 г.)

Fig. 1. Dynamics of OpenAlex replenishment (2010 — September 2025)

ги стран по количеству цитирований в OpenAlex и Scopus остаются высоко коррелированными, хотя в Scopus абсолютные показатели цитирований на страну обычно выше, чем в OpenAlex. Однако для отдельных научных направлений и стран OpenAlex охватывает больше цитирований, чем Scopus и WoS, фиксирует больше идентификаторов ORCID, но меньше рефератов и примерно такое же количество информации открытого доступа [8]. Несмотря на большую степень инклюзивности и широту индексации, этот ресурс сохраняет структурную асимметрию: некоторые регионы и развивающиеся страны представлены в нем менее пропорционально, чем в WoS и Scopus [27].

OpenAlex предоставляет данные в формате JSON<sup>10</sup>, предназначенном для передачи структурированных данных в виде пар ключ — значение (ключ заключается в двойные кавычки, а значение может быть в виде строки, числа или другого объекта) и массивов (упорядоченные списки значений в квадратных скобках).

#### Пример структуры (JSON-like)

```
{
  "title": "Machine Learning in
  Healthcare: A Review",
  "authors": [
    {"name": "John Doe", "affiliation":
    "Harvard University"},
    {"name": "Jane Smith",
    "affiliation": "MIT"}
  ],
  "doi": "10.1234/abcd123",
  "publication date": "2023-05-15",
  "journal": "Journal of Medical AI",
  "abstract": "This paper reviews...",
  "keywords": ["AI", "healthcare",
  "machine learning"],
  "citations": 42,
  "license": "CC BY 4.0",
  "full_text_url": "https://example.
  com/paper.pdf"
}
```

Поиск в огромном массиве документов обеспечивают более 50 поисковых полей OpenAlex.

<sup>10</sup> JSON (JavaScript Object Notation). URL: <https://www.json.org> (accessed: 03.12.2025).

Поисковые фильтры в OpenAlex:

– на английском (51) — Abstract Available, APC Paid (est), Author, Authors Count, Citation Count, Citation Percentile (by Year/subfield), Citation Percentile (year), Continent, Corresponding Author, Corresponding Institution, Countries Count, Country, Cwts Core Source<sup>11</sup>, DOI Prefix, Domain, Field, From Global South, Fulltext, Funder, FWCI, Has Repository Fulltext, In OA Source, Indexed by CrossRef, Indexed by DOAJ, Indexed by MAG Only<sup>12</sup>, Indexed by ORCID, Indexed by Pubmed, Institution, Institution Type, Institutions Count, Keyword, Language, License, Open Access, Open Access Accepted, Open Access Published, Open Access Status, Publisher, Raw Affiliation String, References Count, Related to, Retracted, Source, Source Type, Subfield, Sustainable Development Goal, Title, Title & Abstract, Topic, Type, Year;

– на русском (51) — Аннотация доступна, APC оплачено (оцен.), Автор, Количество авторов, Количество цитирований, Процент цитирования (по году/подполю), Процент цитирования (год), Континент, Автор-корреспондент, Соответствующее учреждение, Количество стран, Страна, Источник CWTS, Префикс DOI, Домен, Поле, Страна Глобального Юга, Полный текст, Финансирующая сторона, FWCI, Имеет репозиторий Fulltext, В источнике открытого доступа, Индексировано CrossRef, Индексировано DOAJ, Индексировано только MAG, Индексировано по ORCID, Индексировано Pubmed, Учреждение, Тип учреждения, Количество учреждений, Ключевое слово, Язык, Лицензия, Открытый доступ, Открытый доступ\_Принят, Открытый доступ\_Опубликовано, Статус открытого доступа, Издатель, Необработанная строка принадлежности, Количество ссылок, Связано с..., Отозвана, Источник, Тип источника, Подполе, Цель устойчивого развития, Название, Заглавие и аннотация, Тема, Тип, Год.

Общее количество полей метаданных в OpenAlex значительно превышает число поисковых полей. Эти поля становятся доступны при выгрузке результатов в виде электронной таблицы (.csv), формате концевой сноски (.ris) или текста (.txt), что позволяет эффективно анализировать большие массивы документов, проводить библиометрический анализ и сохранять данные в удобном

<sup>11</sup> Центр исследований науки и технологий (гол. Centrum voor Wetenschap en Technologies Studies, CWTS) Лейденского университета, Нидерланды, использует различные основные источники для своих исследований и рейтингов, в частности глобального рейтинга университетов CWTS Leiden Ranking и показателей журналов Journal Indicators CWTS.

<sup>12</sup> MAG (Microsoft Academic Graph) — БД компании Microsoft, которая прекратила свое существование 31 декабря 2021 г., но легла в основу OpenAlex.

виде. Часть полей предназначены для сведений об источнике публикации и деталях открытого доступа (например, `is_oa` — наличие открытого доступа, `oa_status` — его тип (золотой, зеленый, гибридный, бронзовый или закрытый), `oa_url` — URL открытой версии). Так, при загрузке метаданных доступны поля ISSN или `is_retracted` (статус отозванной статьи), хотя поисковые поля по этим данным отсутствуют. OpenAlex включает 20 типов описываемых источников, в том числе статьи из журналов, главы книг, препринты, обзоры, наборы данных, отчеты, диссертации и др.

Метаданные в OpenAlex отличаются разнообразием. В частности, в них присутствует классификация публикаций в соответствии с 17 целями устойчивого развития ООН, дополнительные поля метаданных для авторов: `id` — уникальный идентификатор (например, <https://openalex.org/A5115557771> — для автора Редькиной Н.С.), `orcid` — ORCID автора (при наличии), `display_name` — имя для отображения, `works_count` — количество публикаций, `cited_by_count` — общее число цитирований, `last_known_institution` — последняя известная аффилиация, `x_concepts` — тематики, с которыми связан автор. В данном ресурсе не представлены метаданные традиционных для библиотек классификаций (УДК, ББК, Десятичная классификация Дьюи<sup>13</sup> и др.), но реализованы альтернативные способы классификации результатов исследований с применением искусственного интеллекта (ИИ): выделяются области знания и подполя, связанные с тематикой работы.

Пример классификации статьи по теме в OpenAlex

**Пример статьи:** Redkina, N.S. Impact of Open Science Infrastructure on the Development of the World Information Resources Market. *Sci. Tech. Inf. Proc.* 2024. 51, 161–172. <https://doi.org/10.3103/S0147688224700096>

**Abstract.** The promotion of open science initiatives contributes to the development of the world information market and the emergence of a variety of open access resources. The search for scientific information and data, on the one hand, is becoming more extensive, while on the other, it is becoming more difficult due to the ever-increasing amount of resources. It seems that different approaches to structuring open access resources will allow information searches to be more productive and less labor-intensive.

<sup>13</sup> Десятичная классификация М. Дьюи (ДКД, англ. Dewey Decimal Classification, DDC).

**Аннотация.** Продвижение инициатив открытой науки способствует развитию мирового информационного рынка и появлению разнообразных ресурсов открытого доступа (РОД). Поиск научной информации и данных, с одной стороны, становится более расширенным, а с другой — затруднительным из-за все возрастающего количества ресурсов. Представляется, что различные подходы к структурированию РОД позволят проводить информационный поиск продуктивнее и менее трудозатратно.

**Тема:** практики управления исследовательскими данными (блок статей посвящен практикам, проблемам и возможностям обмена данными и управления ими в научных исследованиях. Он охватывает такие темы, как открытая наука, управление исследовательскими данными ...)

**Подполе:** информационные системы (академическое изучение систем с особым упором на информацию и дополнительные сети аппаратного и программного обеспечения, которые люди и организации используют для сбора, фильтрации, обработки, создания и т.д.)

**Область:** компьютерные науки

**Область:** физические науки

**Цель устойчивого развития:** промышленность, инновации и инфраструктура

Подполя определяются путем выявления основной области знания (концепта), к которой относится работа, и последующего сопоставления с наиболее релевантным подполем. Для определения области знания анализируются метаданные статьи и сопоставляются с понятиями из БД Wikidata (например, «информационная инфраструктура», <https://www.wikidata.org/wiki/Q1047251>). Учитываются все совпадения с оценкой релевантности, превышающей 0,3, хотя для некоторых областей знания пороговое значение может быть снижено до 0,1. Считается, что такой подход обеспечивает более высокую полноту охвата, но может снижать точность, включая публикации, которые не всегда соответствуют ожиданиям пользователей.

Пример классификации статьи по области знания из Wikidata в OpenAlex

**Пример статьи:** Redkina, N.S. Impact of Open Science Infrastructure on the Development of the World Information Resources Market *Sci. Tech. Inf. Proc.* 2024. 51, 161–172. <https://doi.org/10.3103/S0147688224700096>

**concepts.id** (идентификатор области знания):

<https://openalex.org/C31652620>|<https://openalex.org/C144133560>

C41008148|<https://openalex.org/C56739046>|<https://openalex.org/C2522767166>|<https://openalex.org/C180198813>|<https://openalex.org/C17744445>|<https://openalex.org/C199539241>

**concepts.wikidata** (ссылка на область знания в Викиданных): <https://www.wikidata.org/wiki/Q1047251>|<https://www.wikidata.org/wiki/Q4830453>|<https://www.wikidata.org/wiki/Q21198>|<https://www.wikidata.org/wiki/Q192060>|<https://www.wikidata.org/wiki/Q2374463>|<https://www.wikidata.org/wiki/Q121182>|<https://www.wikidata.org/wiki/Q36442>|<https://www.wikidata.org/wiki/Q7748>

**concepts.display\_name** (название области знания):

Information infrastructure|Business|Computer science|Knowledge management|Data science|Information system|Political science|Law

**Перевод:** Информационная инфраструктура|Бизнес|Информатика|Управление знаниями|Наука о данных|Информационная система|Политология|Закон

**concepts.score** (оценка релевантности области знания): 0.5071573|0.47695103|0.43841675|0.40247536|0.3793162|0.2539882|0.12366563|0.0

Анализ статей и продемонстрированный выше пример показали, что тематическая классификация статьи часто соответствует содержанию работы, как и некоторые категории областей знания (например, «Информационная инфраструктура» и «Управление знаниями» с оценкой релевантности более 0,3). Однако другие элементы с высокой оценкой, например «Бизнес» (0.47695103), могут быть менее точными. Неточности в классификации могут быть обусловлены отсутствием полного текста, а также недостаточно качественными метаданными, аннотацией или рефератом, которые в неполной мере отражают содержание работы.

Метаданные OpenAlex могут использоваться для построения сетей соавторства, цитирования (подсчет связанных публикаций, совместно цитируемых другими авторами), библиографического сочетания (подсчет публикаций, ссылающихся на одну и ту же работу) с помощью таких специализированных программных инструментов для создания, визуализации и анализа карт, как VOSviewer (<https://www.vosviewer.com/>). Загрузка данных возможна разными способами, включая API. В результате сравнительного анализа API восьми источников данных (Scopus, WoS, Dimensions, OpenAlex, Semantic Scholar, Google Scholar, Scite.ai и Open Research Knowledge Graph) А. Имран и М.К. Паста (A. Imran, M.Q. Pasta) [22]

пришли к выводу, что OpenAlex и Semantic Scholar предлагают наиболее удобный пользовательский интерфейс с минимальными техническими требованиями. Преимуществом OpenAlex является возможность выгрузки библиографических данных миллионов научных статей [16], что позволяет эффективно анализировать и оценивать научные направления по разным параметрам, определять их перспективность и концентрировать усилия на наиболее актуальной тематике. Визуализация данных в виде кластеров трендовых ключевых слов и структурированных библиометрических показателей способствует еще более точному определению приоритетных направлений исследований.

### Dimensions

Dimensions (<https://app.dimensions.ai/>) — известная международная информационно-аналитическая платформа, объединяющая различные виды документов (см. раздел «Материалы»). Анализируя ряд ПОД (Crossref, Dimensions, Microsoft Academic, OpenAlex, Scilit, Semantic Scholar, The Lens), Д. Кьюирос и др. [17] определили, что Dimensions — это продукт с наивысшими показателями полноты метаданных в таких категориях, как рефераты, документы ОД, библиографические данные и типы документов.

Метаданные в Dimensions представлены множеством полей, которые описывают публикации, гранты, клинические испытания, патенты и другие исследовательские объекты, включают общую информацию о публикации, сведения об авторах и аффилиациях, системы классификации и категоризации исследований, такие как Fields of Research (FoR) на основе ANZSRC<sup>14</sup>, RCDC (The Research, Condition, and Disease Categorization) Categories — категории NIH (National Institutes of Health) для биомедицинских исследований, MeSH Terms — медицинские тезаурусные термины, Keywords — ключевые слова. Кроме того, метаданные содержат данные о цитированиях и иных метриках, в том числе количество цитирований (Citation Count), показатель внимания в соцсетях и СМИ (Altmetric Score), индикаторы высокого влияния (Dimensions Badges) и др.

<sup>14</sup> Australian and New Zealand Standard Research Classification (Австралийско-новозеландская стандартная научная классификация) представляет собой набор из трех взаимосвязанных классификаций, разработанных для использования при измерении и анализе статистических данных об исследованиях и разработках (НИОКР) в Австралии и Новой Зеландии (URL: <https://www.arc.gov.au/manage-your-grant/classification-codes-rfcd-seo-and-anzsic-codes>).

### Основные поля метаданных статей в Dimensions

1. Общая информация о публикации:
  - ID — уникальный идентификатор публикации
- в Dimensions,
  - DOI — Digital Object Identifier,
  - PMID — PubMed ID,
  - PMCID — PubMed Central ID,
  - Title — заглавие публикации,
  - Abstract — аннотация,
  - Publication Year — год публикации,
  - Publication Date — дата публикации,
  - Journal Title — название журнала,
  - ISSN/eISSN — идентификаторы журнала,
  - Volume/Issue — том и номер журнала,
  - Pages — страницы публикации,
  - Publisher — издатель,
  - Open Access Status — статус открытого доступа (Gold, Green, Hybrid, Closed),
  - License — лицензия (CC BY, CC BY-NC и др.).
2. Авторы и аффилиации:
  - Authors — список авторов (имя, фамилия, ORCID),
  - Affiliation — институциональная принадлежность авторов (название организации, город, страна),
  - Corresponding Author — контактный автор.
3. Классификация и тематика:
  - Fields of Research (FoR) — категории исследований (ANZSRC),
  - RCDC Categories — категории NIH (для биомедицинских исследований),
  - MeSH Terms — медицинские тезаурусные термины,
  - Keywords — ключевые слова.
4. Цитирования и метрики:
  - Citation Count — количество цитирований,
  - Altmetric Score — показатель внимания в соцсетях и СМИ,
  - Dimensions Badges — индикаторы высокого влияния.
5. Гранты и финансирование:
  - Funding — информация о грантах (название, код, финансирующая организация),
  - Funder — спонсоры (например, NIH, NSF, Wellcome Trust).
6. Связанные объекты:
  - Clinical Trials — связанные клинические испытания,
  - Patents — связанные патенты,
  - Datasets — связанные наборы данных.
7. Дополнительные поля:
  - Language — язык публикации
  - Document Type — тип документа (статья, обзор, глава книги и т.д.),

ISBN — для книг и глав,  
Conference Name — название материалов конференций.

Несмотря на значительный перечень используемых полей метаданных, поиск в бесплатной версии ведется только по некоторым из них. Так, нет возможности осуществлять поиск по всем типам идентификаторов, доступен лишь поиск по DOI. Однако с помощью фильтров результаты поиска можно уточнить по различным критериям, таким как год, тип открытого доступа (закрытый, открытый, зеленый и др.), название источника, вид публикации и др.

Особое значение в Dimensions имеют альтметрические показатели в виде значков, которые отражают внимание к публикациям в социальных сетях (LinkedIn, Weibo и др.), новостных выпусках (например, BBC, The Guardian, Nature News), блогах, патентах и других неакадемических источниках (Википедия, Youtube и др.). К таким показателям относятся Altmetric Attention Score — агрегированный показатель внимания, рассчитываемый на основе упоминаний в разных источниках, Mentions Count — общее количество упоминаний публикации и Sources Count — количество уникальных источников, упоминающих публикацию. В данных Dimensions также можно найти количество упоминаний в каждом источнике (например, 50 твитов, 3 новостные статьи), географию упоминаний (страны, где обсуждается публикация), динамику упоминаний (как меняется внимание со временем). Определенными значками отмечаются высокоцитируемые публикации или быстро набирающие внимание, а также публикации открытого доступа. В бесплатной версии Dimensions возможен экспорт не более 500 записей в формате XLSX-Excel для дальнейшего библиометрического анализа с помощью VOSviewer или CiteSpace, для менеджера ссылок в BibTeX.

Сравнивая БД OpenAlex и Dimensions, можно сделать следующие выводы: оба ПОД имеют данные о цитированиях, но Dimensions учитывает патенты, цитирования протоколов клинических исследований, а также предоставляет альтметрические показатели внимания в соцсетях и СМИ; OpenAlex использует концепты (concepts) на основе Fields of Study, а Dimensions — области исследования на основе ANZSRC. Метаданные в этих ресурсах помогают анализировать научные публикации, отслеживать тренды, оценивать влияние исследований и находить коллаборации.

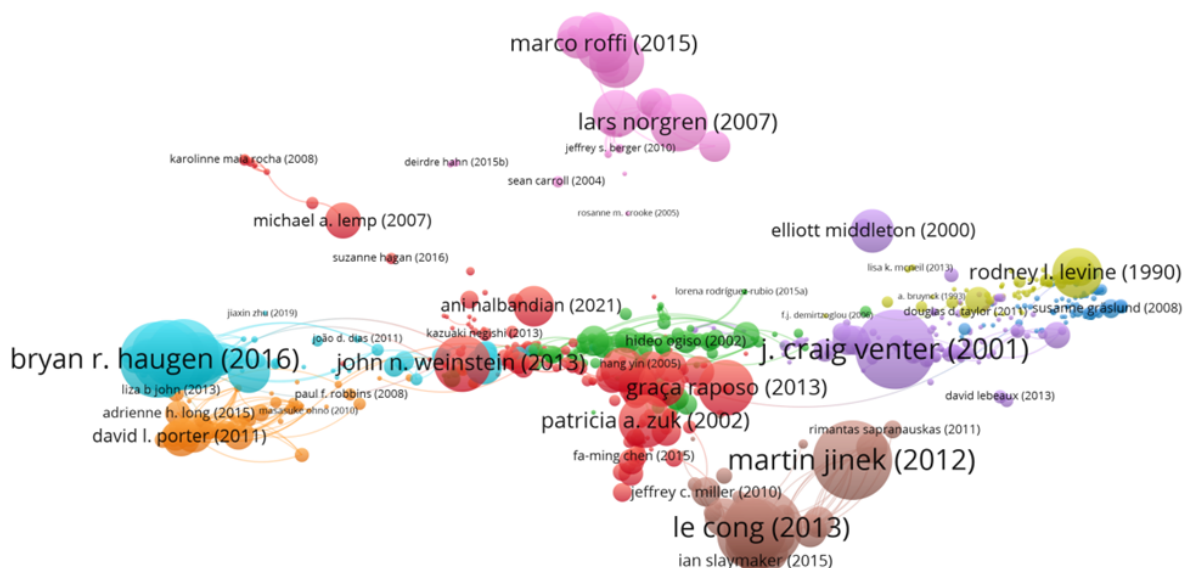
#### *The Lens*

The Lens — это агрегатор научной информации, содержащий более 270 млн документов, собран-

ных и унифицированных из таких источников, как Microsoft Academic Graph (MAG, проект прекратил обновления, однако его данные были импортированы), CORE, PubMed и Crossref, с дополнениями из OpenAlex и браузерного расширения UnPaywall, а также с привязкой к профилям ORCID. Платформа The Lens (<https://www.lens.org/>) агрегирует метаданные научных статей, патентов и исследовательских работ. Для каждой публикации доступны следующие основные метаданные:

- название статьи,
- авторы (с аффилиациями на момент публикации и ссылками на нормализованные профили организаций, с указанием числа публикаций автора в The Lens),
- DOI (Digital Object Identifier),
- дата публикации,
- журнал/конференция (название, ISSN, издатель),
- аннотация,
- ключевые слова,
- библиометрические данные,
- цитирования (количество цитирований с указанием ссылающихся источников),
- информация об индексации (Scopus, PubMed, WoS и др.),
- тип лицензии (например, CC BY 4.0),
- ссылка на полный текст (PDF) во внешнем источнике (при наличии открытого доступа).

The Lens решает главные проблемы метаданных, связанные с их несогласованностью, объединяя разные варианты написания имени одного автора в один профиль, привязывая публикации к конкретным организациям с помощью открытых идентификаторов (GRID/ROR) и нормализуя названия журналов, конференций. The Lens указывает все постоянные идентификаторы: DOI, PMID (PubMed), PMCID (PubMed Central), ISSN, ISBN и др., что позволяет точно идентифицировать документ и без ошибок переходить к полному тексту. В анализируемом ПОД приводятся библиометрические данные, причем списки цитирующих и цитируемых работ представлены непосредственно в интерфейсе. Уникальной особенностью The Lens являются ссылки на патенты, которые цитируют эту статью, и наоборот. Поиск в The Lens поддерживает фильтрацию по концептам (Concepts) — автоматически присваиваемым темам / ключевым словам из большой иерархической онтологии, а также по специализированным предметным классификациям (если они представлены в метаданных): MeSH (для медицины), ACM (Association for Computing Machinery — Ассоциация вычислительной техники), PACS (Physics and Astronomy Classification



**Рис. 2.** Сеть библиографических связей на основе цитирований российских патентов, построенная в VOSviewer по данным The Lens (октябрь 2025)

**Fig. 2.** Visualization of bibliographic coupling in citations of Russian patents in VOSviewer (based on The Lens data, September 2025)

Scheme — Система классификации для физики и астрономии) и др.

Метаданные The Lens позволяют визуализировать сети соавторства и сотрудничества организаций, выявлять тренды по темам (концептам), анализировать патентно-публикационную активность (рис. 2).

В VOSviewer метаданные The Lens могут быть загружены как напрямую в виде файла, так и с помощью менеджеров ссылок (EndNote, RefWorks). Кроме того, часть метаданных доступна через открытый API, что позволяет ученым самим проводить аналитические исследования.

Пример поискового запроса для API The Lens:

```
{
  "query": {
    "bool": {
      "must": [
        {"match": {"authors.name": "Redkina Natalya"}},
        {"match": {"affiliations.city": "Novosibirsk"}}
      ]
    }
  },
  "size": 10
}
```

Как и в выше проанализированных РОД, качество данных в The Lens напрямую зависит от качества исходных метаданных, загружаемых из Crossref, PubMed и других источников. На эту платформу могут переноситься ошибки в аффилиациях или именах авторов, присутствующие в первоисточнике. Кроме того, показатели цитирования, рассчитываемые этой платформой, не являются общепринятой метрикой. Однако они могут быть полезны для внутреннего анализа.

### Semantic Scholar

Semantic Scholar — это бесплатная платформа для поиска научных публикаций, предоставляющая доступ к более чем 230 млн публикаций по всем областям знаний. Корпус документов сформирован благодаря партнерству с издателями (Wiley, Cambridge University Press и др.), поставщиками и агрегаторами данных (Crossref, PubMed, Unpaywall) и собственному веб-сканеру. На платформе используются методы извлечения научного контента из PDF-файлов, и на этой основе автоматически строится граф знаний Semantic Scholar Academic Graph, содержащий, помимо научной литературы, метаданные более 100 млн авторов, более 650 млн связей между статьями и авторством, более 2,8 млрд связей цитирования. До прекраще-

ния поддержки MAG компания Semantic Scholar использовала ее классификацию областей исследований, после перешла на собственную модель классификации, в которую были добавлены такие области, как образование, право и лингвистика.

Результаты поиска можно фильтровать с помощью соответствующих фасетов по области научных знаний, периоду охвата, автору, доступности полного текста и журналу, в котором была опубликована статья. Ранжировать документы можно по их релевантности, количеству цитирований и влиятельности или по новизне. Встроенные алгоритмы ИИ позволяют генерировать краткое резюме по статье. В Semantic Scholar предоставляется возможность перекрестного поиска, т.е. отображаются работы, цитируемые автором рассматриваемой статьи, а также работы авторов, которые позднее ссылались на эту статью. Таким образом, отслеживая цепочку цитирований, пользователь может проследить преемственность идей в научном направлении. ИИ предлагает просмотр связанных работ в отдельной вкладке. В профиле ученого представлены ORCID и наиболее значимые количественные показатели работы по версии Semantic Scholar: количество статей, индекс Хирша, общее количество цитирований и количество наиболее влиятельных цитирований. Особенность данного РОД — автор может добавлять статьи в свой профиль.

Еще одним преимуществом Semantic Scholar является дополнение часто ограниченной информации о статье данными, извлеченными из PDF-файла (при наличии доступа к полному тексту). В результате становятся доступными библиографические данные, на основе которых строится граф цитирования, а также структурированное содержание: заголовки разделов, абзацы, рисунки и таблицы, встроенные ссылки на них и на библиографические записи. Извлечение из PDF-файлов структурированной информации является сложной задачей и может сопровождаться ошибками. Тем не менее это дает более полное представление о статье.

Для неавторизованных пользователей предлагается небольшой объем API-запросов и тестовых наборов данных. Для получения полных наборов данных и увеличения лимита запросов необходимо бесплатно получить ключ доступа, который предоставляется после согласия с условиями использования (<https://www.semanticscholars.org/product/api##Partner-Form>). Для получения дополнительной информации о работе с API, включая форматирование запросов и анализ ответов, необходимо обратиться к официальной документации (<https://api.semanticscholar.org/api-docs/>).

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты подтверждают, что крупнейшие РОД по объему охвата публикаций сопоставимы с коммерческими наукометрическими БД. Вместе с тем качество представления метаданных в этих ресурсах остается неоднородным: отдельные поля (аффилиация, идентификаторы авторов (ORCID), сведения о финансировании, библиографические ссылки, аннотации, ключевые слова и тип документа) существенно различаются по полноте, согласованности и надежности. Таким образом, РОД могут рассматриваться как значимый альтернативный источник научной информации, однако их использование в аналитических и оценочных исследованиях требует методологической осторожности. Эти наблюдения согласуются с выводами в работах Л. Делгадо-Кирос и Дж. Л. Ортеги [6], Н.А. Мазова и В.Н. Гуреева [10, 11], которые установили, что Dimensions, OpenAlex, Scilit и The Lens обеспечивают большую полноту метаданных по сравнению с научными поисковыми системами<sup>15</sup>, хотя между самими ресурсами сохраняются различия по структуре и качеству метаданных.

Результаты настоящего исследования согласуются также с выводами Дж. Кулберта с соавт. [8], показавшими, что OpenAlex уступает коммерческим наукометрическим БД по точности тематической классификации и по степени стандартизации метаданных и корректности управления ими. Рост охвата публикаций на этой платформе не сопровождается пропорциональным ростом надежности их метаданных. Выявленные различия между OpenAlex и Dimensions в сведениях о финансировании подтверждают выводы Е. Шареса [12]. Зависимость OpenAlex от Crossref как основного источника метаданных ограничивает полноту информации о грантах, тогда как Dimensions расширяет эти данные за счет парсинга разделов «Благодарности» и соглашений с издателями. Таким образом, различия в полноте сведений о финансировании обусловлены прежде всего архитектурой источников пополнения метаданных, а не только функциональными возможностями самих платформ.

Проблемы тематической классификации, выявленные в OpenAlex при сопоставлении с концептами Wikidata, соответствуют выводам Л. Сеспедес

<sup>15</sup> Под научными поисковыми системами в данном исследовании понимаются агрегаторы научных публикаций, ориентированные на поиск и навигацию (например, Google Scholar, Semantic Scholar, BASE), в отличие от структурированных библиографических БД открытого доступа, таких как Dimensions, OpenAlex, Scilit и The Lens.

и соавт. [7] о снижении точности алгоритмической тематической атрибуции при высокой полноте охвата. Алгоритмические классификации обеспечивают эффективное выявление крупных тематических кластеров, однако не гарантируют корректной дисциплинарной стратификации отдельных публикаций. Использование таких классификаций целесообразно ограничивать задачами макроаналитического уровня.

Полученные результаты согласуются также с результатами исследования П. Сингх с соавт. [13], которые показали, что, хотя OpenAlex и Dimensions и обладают широкими аналитическими возможностями и охватом исследовательских тем, они остаются ограниченно пригодными для оценочной наукометрии из-за проблем с полнотой и надежностью метаданных. Аналогичную позицию занимают А. Велес-Эстевес с соавт. [14], указывая на более низкое качество метаданных OpenAlex по ряду ключевых полей в сравнении с Scopus и WoS, особенно для публикаций, индексируемых исключительно в открытых ресурсах.

Функциональные различия между анализируемыми ресурсами формируют различные сценарии их практического использования. Dimensions предоставляет развитые альтиметрические и аналитические инструменты, позволяющие анализировать внимание к публикациям за пределами академического цитирования. OpenAlex обеспечивает наибольшую открытость и масштабируемость выгрузки данных, что повышает его ценность для воспроизводимых библиометрических исследований. The Lens расширяет аналитические возможности за счет интеграции патентных и публикационных данных. Semantic Scholar ориентирован на семантическое обнаружение связей между публикациями и визуализацию цитатного влияния. Подобная дифференциация подтверждает выводы И.Д. Тургель и О.А. Черновой [9] о комплементарном характере РОД.

Отдельного внимания требует проблема несовместимости классификационных систем. Использование концептов OpenAlex, категорий FoR в Dimensions, онтологий The Lens и тегов Semantic Scholar делает межресурсное тематическое сопоставление методологически затруднительным. Проведение корректного сравнительного анализа требует строгой фиксации выбранной онтологии и отказа от смешения классификационных систем в рамках одного исследования, на что также указывают П. Сингх с соавт. [13] и А. Велес-Эстевес и др. [14].

С практической точки зрения результаты данного исследования подчеркивают критическую роль

качества первичных метаданных. Поддержка ORCID, нормализация аффилиаций по ROR, полнота аннотаций, корректное указание ключевых слов и сведений о финансировании напрямую определяют качество последующего индексирования публикаций в РОД. Эти выводы соответствуют целям инициатив Metadata 20/20, I4OC, I4OA и COMET, направленных на повышение полноты, связности и повторного использования открытых метаданных.

Для исследователей и библиотечных специалистов полученные результаты означают необходимость осознанного выбора ресурсов в зависимости от целей исследования, а также обязательного документирования процедур очистки и нормализации метаданных при проведении библиометрических и систематических обзоров. Как показали У.Х. Уолтерс [5] и М. Гузенбауэр [15], выбор БД напрямую влияет на полноту выявления публикаций и ссылок, что делает комбинированное использование нескольких источников методологически оправданным.

Совокупность полученных результатов подтверждает, что РОД формируют новую инфраструктуру научной коммуникации, основанную на принципах открытости, связности и масштабируемости анализа. Применение таких ресурсов в оценочной и аналитической наукометрии требует развитой методологической культуры интерпретации метаданных и строгого учета ограничений каждого ресурса.

## ОГРАНИЧЕНИЯ ИССЛЕДОВАНИЯ

Настоящее исследование направлено на выявление и анализ научных РОД, позволяющих не только осуществлять многопараметрический поиск различных документов, но и выгружать значительные объемы метаданных для дальнейших библиометрических исследований и визуализации. Оно имеет определенные ограничения как по количеству выбранных РОД, так и по методологии. Фокус внимания сосредоточен на четырех научных РОД (OpenAlex, Dimensions, The Lens и Semantic Scholar) с датировкой анализа (октябрь 2025 г.), разным объемом и видовым составом включенных документов и, соответственно, набором полей метаданных, источниками и технологиями пополнения, собственными подходами к выбору классификации документов, различиями в форматах и объемах выгрузки документов, возможностями проведения дальнейшего библиометрического анализа и визуализации полученных данных с помощью открытых инструментов. Такой исследовательский дизайн был избран с целью очертить контуры дальнейшего анализа

возможностей и проблем метаданных, отражаемых в иных научных РОД, представленных на мировом информационном рынке (CORE, BASE и др.), и может служить отправной точкой для будущих исследований, реализованных в рамках аналитического и сравнительного подходов представления полей метаданных в научных информационных ресурсах и их сервисных возможностях.

## ЗАКЛЮЧЕНИЕ

Проведенное исследование позволило обобщить представления о функционировании крупнейших научных РОД и оценить их значение для информационного поиска и аналитической обработки научных публикаций. Зафиксировано, что РОД формируют самостоятельный и быстро развивающийся сегмент научной информационной инфраструктуры, обладающий значительным потенциалом для поиска, анализа и визуализации научных данных.

Ключевой вывод исследования состоит в том, что эффективность РОД определяется не только масштабом охвата публикаций, но прежде всего качеством, структурной согласованностью и связностью их метаданных. В данном контексте метаданные выступают не вспомогательным элементом описания, а базовым инструментом управления научной информацией, влияющим на обнаруживаемость публикаций, корректность аналитических выводов и интерпретируемость результатов библиометрических исследований. Таким образом, использование РОД требует осознанного методологического подхода и учета ограничений, связанных с неоднородностью и источниками формирования метаданных.

Результаты исследования показывают, что проанализированные РОД обладают различным функциональным профилем и ориентированы на решение разноплановых исследовательских задач. Это обстоятельство подтверждает необходимость их дифференцированного и взаимодополняющего использования в зависимости от целей исследования, а также отказа от универсализации отдельных ресурсов в качестве единственного источника аналитических данных. В этом смысле РОД сле-

дует рассматривать не как замену коммерческих наукометрических БД, а как самостоятельную инфраструктуру, расширяющую возможности научного анализа при соблюдении методологических условий.

Практическая значимость полученных результатов связана с необходимостью повышения культуры работы с метаданными на всех уровнях научной коммуникации. Для редакций журналов ключевым становится обеспечение полноты и корректности метаданных, передаваемых в Crossref, включая сведения об аффилиациях, финансировании, аннотациях и использовании международных идентификаторов, а также регулярный аудит соответствия DOI, лицензий и статусов открытого доступа. Для авторов существенное значение имеет поддержка актуального профиля ORCID, контроль корректного отображения публикаций в крупнейших РОД и осознанный выбор ключевых слов с учетом используемых онтологий и классификаций. Для библиотек и разработчиков информационных систем результаты исследования подчеркивают важность выстраивания воспроизводимых процедур загрузки и обработки метаданных, включающих документирование этапов очистки, нормализации и валидации данных до проведения библиометрического анализа и визуализации. В совокупности эти меры создают условия для повышения надежности аналитических выводов и интеграции РОД в устойчивую инфраструктуру научных исследований.

В более широком контексте исследование подтверждает, что развитие РОД отражает системный переход от замкнутых наукометрических платформ к распределенной инфраструктуре открытых научных данных. В этой инфраструктуре метаданные становятся ключевым связующим элементом между публикациями, исследователями, организациями и аналитическими инструментами. Дальнейшее развитие данной экосистемы связано с совершенствованием стандартов метаданных, расширением практик их совместного обогащения и усилением роли библиотек и научных журналов как активных участников процессов научной коммуникации в условиях открытой науки.

## ФИНАНСИРОВАНИЕ

Исследование выполнено в рамках реализации научного проекта ГПНТБ СО РАН (2022-2026 гг.) «Разработка модели функционирования научной библиотеки в информационной экосистеме открытой науки», № 122041100150-3.

## FUNDING

The study was carried out within the framework of the scientific project of the State Public Scientific and Technological Library of the Siberian Branch of the Russian Academy of Sciences (2022-2026) «Development of a Model for the Functioning of a Scientific Library in the Information Ecosystem of Open Science», № 122041100150-3.

**КОНФЛИКТ ИНТЕРЕСОВ**

Автор заявляет об отсутствии конфликта интересов.

**CONFLICT OF INTERESTS**

The author declares no relevant conflict of interests.

**СПИСОК ЛИТЕРАТУРЫ / REFERENCES**

1. Elouataoui W., El Alaoui I., Gahi Y. Metadata quality dimensions for Big Data use cases. In: *Proc. 2<sup>nd</sup> Int. Conf. on Big Data, Modelling and Machine Learning*. Setúbal: SciTePress; 2021:488-495. <https://doi.org/10.5220/0010737400003101>
2. Salvesen L.M. The future of enriched, linked, open and filtered metadata making sense of IFLA LRM, RDA, linked data and BIBFRAME. (2022), by Getaneh Alemu, London, UK: Facet Publishing, 222 pp. ISBN 9781783304943. *Journal of Electronic Resources Librarianship*. 2024;36(4):336-337. <https://doi.org/10.1080/1941126X.2024.2417127>
3. Frederick D.E. Are library data and metadata dying or nearly dead? A discussion for everyone. *Library Hi Tech News*. 2024;41(10):1-12. <https://doi.org/10.1108/LHTN-08-2024-0148>
4. Roy N., Gayan M.A. Metadata quality of selected South Asian national libraries' Web OPACs: An assessment of cataloging errors. *Cataloging & Classification Quarterly*. 2025;63(1):57-83. <https://doi.org/10.1080/01639374.2025.2451158>
5. Walters W.H. Comparing conventional and alternative mechanisms of discovering and accessing the scientific literature. *Proceedings of the National Academy of Sciences of the United States of America*. 2025;122(27):e2503051122. <https://doi.org/10.1073/pnas.2503051122>
6. Delgado-Quirós L., Ortega J.L. Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies*. 2024;5(1):31-49. [https://doi.org/10.1162/qss\\_a\\_00286](https://doi.org/10.1162/qss_a_00286)
7. Céspedes L., Kozłowski D., Pradier C., et al. Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*. 2025;76(6):884-895. <https://doi.org/10.1002/asi.24979>
8. Culbert J., Hobert A., Jahn N., et al. Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*. 2025;130(4):2475-2492. <https://doi.org/10.1007/s11192-025-05293-3>
9. Turgel I.D., Chernova O.A. Open science alternatives to Scopus and the Web of Science: A case study in regional resilience. *Publications*. 2024;12(4):43. <https://doi.org/10.3390/publications12040043>
10. Мазов Н.А., Гуреев В.Н. Ведение базы данных публикаций организации с использованием библиографических ресурсов открытого доступа. *Научно-техническая информация. Серия 1: Организация и методика информационной работы*. 2023;(9):20-32. <https://doi.org/10.36535/0548-0019-2023-09-4>  
Mazov N.A., Gureyev V.N. Maintaining a database of an organization's publications using open access bibliographic resources. *Nauchno-tekhnicheskaya informatsiya. Seriya 1: Organizatsiya i metodika informatsionnoy raboty*. 2023;(9):20-32. (In Russ.). <https://doi.org/10.36535/0548-0019-2023-09-4>
11. Mazov N.A. Gureyev V.N. Open access bibliographic resources for maintaining a bibliographic database of research organization. *Scientific and Technical Information Processing*. 2023;50(3): 211-223. <https://doi.org/10.3103/S0147688223030115>
12. Schares E. Comparing funder metadata in OpenAlex and Dimensions. OpenIS U. April 23, 2024. <https://doi.org/10.31274/b8136f97.ccc3dae4>
13. Singh P., Singh V.K., Kanaujia A. Metadata fields in Web of Science, Scopus, Dimensions and OpenAlex Databases: An exploratory analysis of the possibilities and ease of doing scientometric analyses. In: *Proc. 5<sup>th</sup> Int. Conf. on Science & Technology Metrics (STMet 2024)*. My Tho: Tien Giang University; 2024:214-222. <https://doi.org/10.6025/stm/2024/5/214-222>
14. Velez-Estevez A., Perez I.J., García-Sánchez P., Moral-Munoz J.A., Cobo M.J. New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*. 2023;60(4):103385. <https://doi.org/10.1016/j.ipm.2023.103385>
15. Gusenbauer M. Beyond Google Scholar, Scopus, and Web of Science: An evaluation of the backward and forward citation coverage of 59 databases' citation indices. *Research Synthesis Methods*. 2024;15(5):802-817. <https://doi.org/10.1002/jrsm.1729>
16. Редькина Н.С. Ресурсы открытого доступа и инструменты для определения трендов развития науки. *Научные и технические библиотеки*. 2025;(4):90-113. <https://doi.org/10.33186/1027-3689-2025-4-90-113>

- Redkina N. S. Open-access resources and tools to identify the trends of science advancement. *Scientific and Technical Libraries*. 2025;(4):90-113. (In Russ.). <https://doi.org/10.33186/1027-3689-2025-4-90-113>
17. Quirós L.J.D., Ortega J.L. Comparing bibliographic descriptions in seven free-access databases. In: *Proc. 27<sup>th</sup> Int. Conf. on Science, Technology and Innovation Indicators (STI 2023)*. (Leiden, September 27-29, 2023). Leiden: Centre for Science and Technology Studies; 2023. <https://doi.org/10.55835/6436c590b3340c364be5b2c7>
  18. Alperin J.P., Portenoy J., Demes K., et al. An analysis of the suitability of OpenAlex for bibliometric analyses. arXiv preprint. 2024. <https://doi.org/10.48550/arXiv.2404.17663>
  19. Priem J., Piwovar H., Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. In: *From Global Indicators to Local Applications. Proc. 26<sup>th</sup> Int. Conf. on Science, Technology and Innovation Indicators (STI 2022)*. (Granada, September 7-9, 2022). Leiden: Centre for Science and Technology Studies; 2022. <https://doi.org/10.48550/arXiv.2205.01833>
  20. Haunschild R., Bornmann L. Usage of OpenAlex for creating meaningful global overlay maps of science on the individual and institutional levels. arXiv preprint. 2024. <https://doi.org/10.1371/journal.pone.0308041>
  21. Akbaritabar A., Theile T., Zagheni E. Global flows and rates of international migration of scholars. *MPIDR Working Paper*. 2023;(18). <https://doi.org/10.4054/MPIDR-WP-2023-018>
  22. Imran A., Pasta M. Q. A qualitative study of bibliographic data sources and retrieval mechanisms for computational research. In: *Proc. 26<sup>th</sup> Int. Multi-Topic Conf. (INMIC)*. (Karachi, December 30-31, 2024). New York, NY: IEEE; 2024:1-6. <https://doi.org/10.1109/INMIC64792.2024.11004418>
  23. Zhang L., Cao Z., Shang Y., et al. Missing institutions in OpenAlex: Possible reasons, implications, and solutions. *Scientometrics*. 2024;129(10):5869-5891. <https://doi.org/10.1007/s11192-023-04923-y>
  24. Scheidsteger T., Haunschild R., Hug S., et al. The concordance of field-normalized scores based on Web of Science and Microsoft Academic data: A case study in computer sciences. In: *Science, Technology and Innovation Indicators in Transition. Proc. 23<sup>rd</sup> Int. Conference on Science, Technology and Innovation Indicators (STI 2018)*. (Leiden, September 12-14, 2018). Leiden: Centre for Science and Technology Studies; 2018. URL: [https://www.researchgate.net/publication/324978237\\_The\\_concordance\\_of\\_field-normalized\\_scores\\_based\\_on\\_Web\\_of\\_Science\\_and\\_Microsoft\\_Academic\\_data\\_A\\_case\\_study\\_in\\_computer\\_sciences](https://www.researchgate.net/publication/324978237_The_concordance_of_field-normalized_scores_based_on_Web_of_Science_and_Microsoft_Academic_data_A_case_study_in_computer_sciences)
  25. Mezquita B. A comparison of OpenAlex with Scopus and Web of Science for tracking scholarly nursing literature. *SAGE Open Nursing*. 2025;11:23779608251361012. <https://doi.org/10.1177/23779608251361012>
  26. Rajit D., et al. Assessing the coverage of PubMed, Embase, OpenAlex, and Semantic Scholar for automated single-database searches in living guideline evidence surveillance: A case study of the international polycystic ovary syndrome guidelines 2023. *Journal of Clinical Epidemiology*. 2025;183:111789. <https://doi.org/10.1016/j.jclinepi.2025.111789>
  27. Maddi A., Maisonobe M., Boukacem-Zeghmouri C. Geographical and disciplinary coverage of open access journals: OpenAlex, Scopus, and WoS. *PLoS One*. 2025;20(4):e0320347. <https://doi.org/10.1371/journal.pone.0320347>

### ИНФОРМАЦИЯ ОБ АВТОРЕ

**Наталья Степановна Редькина**, доктор педагогических наук, заведующий отделом, отдел научных исследований открытой науки Государственной публичной научно-технической библиотеки Сибирского отделения Российской академии наук (ГПНТБ СО РАН), г. Новосибирск, Российская Федерация; <https://orcid.org/0000-0002-3486-9711>; e-mail: [redkina@spsl.nsc.ru](mailto:redkina@spsl.nsc.ru)

### INFORMATION ABOUT THE AUTHOR

**Natalya S. Redkina**, Dr. Sci. (Educ.), Head of Department of the Open Science Research Department, State Public Scientific Technological Library, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation; <https://orcid.org/0000-0002-3486-9711>; e-mail: [redkina@spsl.nsc.ru](mailto:redkina@spsl.nsc.ru)

Поступила в редакцию / Received 03.10.2025

Поступила после рецензирования / Revised 05.11.2025

Принята к публикации / Accepted 26.12.2025