



DOI: 10.19181/sntp.2024.6.4.5

EDN: IQCJOK

Научная статья

Research article

РОССИЙСКАЯ НАУКА В СИСТЕМЕ ОТКРЫТЫХ НАУЧНЫХ ЗНАНИЙ OPENALEX



**Редькина
Наталья Степановна¹**

¹ Государственная публичная научно-техническая библиотека СО РАН, Новосибирск, Россия

Для цитирования: Редькина Н. С. Российская наука в системе открытых научных знаний OpenAlex // Управление наукой: теория и практика. 2024. Т. 6, № 4. С. 86–104. DOI 10.19181/sntp.2024.6.4.5. EDN IQCJOK.

Аннотация. OpenAlex – активно развивающийся ресурс открытого доступа, а также многообещающая альтернатива коммерческим источникам библиометрических данных, нацеленная на поддержку открытых исследовательских практик и трансформацию информационной экосистемы мировой науки. Текущее исследование представляет собой анализ общего массива представленных публикаций на платформе OpenAlex, а также более детальное изучение работ российских учёных в данной системе открытых научных знаний. Для достижения поставленной цели был осуществлён сбор информации и её обработка по годам публикаций, видам и наиболее актуальной тематике исследований, издателям, наличию наборов данных, наиболее цитируемым статьям (> 50), их распределению по предметным областям и другим параметрам. Результатом исследования стали готовые визуализированные данные, которые позволили проанализировать публикации российских учёных в сравнении с зарубежными источниками, а также выявить закономерности цитирования, проблемы и перспективы вхождения науки в инфраструктуру открытой науки. Доля публикаций российских учёных от общего массива в OpenAlex по состоянию на 01.07.2024 г. составила не более 1% (2 398 923) с преобладанием в документопотоке научных статей. Однако зафиксировано значительное увеличение числа публикаций в последние два года. Отмечено, что эффективное представление в растущем потоке открытых данных и публикаций возможно посредством популяризации среди исследователей идей открытой науки и интеграции российских ресурсов в мировую информационную систему открытого доступа. Это позволит улучшать видимость, повышать влияние результатов исследования и цитируемость работ, расширять географию читательской аудитории, получать иные преимущества открытого доступа. Сделан вывод, что OpenAlex может быть использован в качестве информационного ресурса для поиска разнородной информации и альтернативного инструмента библиометрического анализа, что особенно актуально для российских исследователей в условиях рестрикций к зарубежным лицензионным базам данных.

Ключевые слова: открытая наука, открытый доступ, информационные ресурсы, поиск информации, библиометрия, OpenAlex

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

RUSSIAN SCIENCE IN THE OPENALEX SYSTEM OF OPEN SCIENTIFIC KNOWLEDGE

Natalya S. Redkina¹

¹ State Public Scientific Technological Library, SB RAS, Novosibirsk, Russia

For citation: Redkina N. S. Russian science in the OpenAlex system of open scientific knowledge. *Science Management: Theory and Practice*. 2024;6(4):86–104. (In Russ.). DOI 10.19181/sntp.2024.6.4.5.

Abstract. OpenAlex is an actively developing open access resource and a promising alternative to commercial sources of bibliometric data. It is aimed at supporting open research practices and transforming the information ecosystem of global science. The current study is an analysis of the total array of publications presented on the OpenAlex platform, as well as a more detailed study of Russian researchers' works indexed in this system of open scientific knowledge. To achieve this goal, information was collected, processed and arranged by years of publication, types and most relevant research topics, publishers, availability of datasets, citation rates (> 50), their subject area distribution and other parameters. The study resulted in obtaining ready-made visualized data that made it possible to analyze publications written by Russian researchers in comparison with foreign sources, as well as to find out citation patterns, problems and prospects for science to enter the open science infrastructure. The share of publications by Russian researchers in the total array as of July 1, 2024 was no more than 1% (2,398,923), with a predominance of scientific articles in the flow. However, a significant increase in the number of publications has been recorded in the last two years. It is noted that an effective representation in the growing flow of open data and publications is possible through the popularization of open science ideas among researchers and the integration of Russian resources into the global open access information system. This will improve visibility, increase the impact of research results and citation rate of works, expand the geography of the readership and allow receiving other benefits of open access. The author concludes that OpenAlex can be used as an information resource to search for heterogeneous information and an alternative tool for bibliometric analysis. This is especially important for Russian researchers in the context of limited access to foreign licensed databases.

Keywords: open science, open access, information resources, information retrieval, bibliometrics, OpenAlex

Acknowledgements. The article was prepared according to the research plan of the State Public Scientific Technological Library of the Siberian Branch of the Russian Academy of Sciences, the project "Development of a Model for the Functioning of a Scientific Library in the Information Ecosystem of Open Science", No. 122041100150-3.

ВВЕДЕНИЕ

Открытая наука – это движение, цели которого были определены основополагающими декларациями ещё в 2002–2003 гг. (Будапештская инициатива открытого доступа и др.), которое уже навряд ли можно остановить. Концепция открытой науки направлена на распространение научных знаний, их общедоступность и многократное использование, она расширяет научное сотрудничество, обмен и оценку результатов исследований, расширяет географические границы, увеличивает видимость и цитируемость научных работ, обеспечивает более быстрое распространение инноваций и даёт другие преимущества [1].

В опубликованной в 2024 г. Барселонской декларации об открытой научной информации говорится, что в то время, когда при принятии решений в науке всё чаще используются аналитические показатели, решение проблем закрытости исследовательской информации должно стать одним из главных приоритетов¹. Результатом развития идей и принципов открытой науки стало распространение ресурсов открытого доступа, предоставляющих широкий спектр информационно-поисковых и библиометрических возможностей. Речь идёт не только о публикациях в открытом доступе, но и о платформах открытых данных, где размещаются богатейшие ресурсы, играющие решающую роль в практике открытых научных знаний [2], (мета) данных [3], рефератов (Инициатива открытости рефератов², призывающая научных издателей открывать рефераты опубликованных работ и, по возможности, отправлять их в Crossref), препринтов³, патентов⁴, рецензий⁵, в том числе на препринты [4], показателей результативности (Инициатива по открытому цитированию⁶, цель которой – создание глобальной публичной сети связанных данных научных цитирований для улучшения возможности обнаружения опубликованного контента как по подписке, так и в открытом доступе) и др.

Международный научный совет (International Science Council), обсуждая будущее научных публикаций, не без основания предположил, что в 2024 г. ожидается активизация усилий по обеспечению открытого доступа к научной литературе и данным исследований, а также отметил растущую тенденцию к использованию баз данных с открытым доступом, таких как Lens и OpenAlex, которые «могут дополнять или предлагать альтернативы коммерческим базам данных, таким как Scopus и Web of Science»⁷. Этой же позиции придерживаются многие исследователи и организации [5; 6]. Французский национальный центр научных исследований (Centre National de la Recherche Scientifique, CNRS),

¹ Barcelona Declaration on Open Research Information : [сайт]. URL: <https://barcelona-declaration.org> (дата обращения: 31.07.2024).

² I4OA: Initiative for Open Abstracts : [сайт]. URL: <https://i4oa.org> (дата обращения: 31.07.2024).

³ PREREVIEW welcomes new funding to support our efforts over the next three years // PREREVIEW : [сайт]. 2024. June 4. URL: <https://content.prereview.org/prereview-welcomes-funding> (дата обращения: 31.07.2024).

⁴ The Principles of Open Scholarly Infrastructure : [сайт]. URL: <https://openscholarlyinfrastructure.org> (дата обращения: 31.07.2024).

⁵ Peer Reviewers' Openness Initiative : [сайт]. URL: <https://opennessinitiative.org> (дата обращения: 31.07.2024).

⁶ I4OC: Initiative for Open Citations : [сайт]. URL: <https://i4oc.org/> (дата обращения: 31.07.2024).

⁷ Предстоящий многообещающий год для научных публикаций // Международный научный совет : [сайт]. 2024. 30 января. URL: <https://ru.council.science/blog/a-promising-year-ahead-for-scientific-publishing/> (дата обращения: 31.07.2024).

активно продвигающий политику открытой науки, в 2024 г. объявил об отказе от дальнейшей подписки на коммерческую библиографическую базу данных Scopus и поддержке перехода на бесплатные библиометрические инструменты: OpenAlex (<https://openalex.org>), Crossref (<https://crossref.org>), Dimensions (<https://dimensions.ai>) и открытый национальный архив HAL (<https://hal.science/>), «использование которых способствует повышению доступности результатов научной деятельности»⁸ (здесь и далее пер. мой. — *Н. Р.*).

Отдельные страны и учреждения активно продвигают инициативы по поддержке и использованию ресурсов и инструментов открытого доступа. Так, Министерство высшего образования и научных исследований Франции (Ministère de l'enseignement supérieur et de la recherche, MESR) установило многолетнее партнёрство с OpenAlex⁹. Партнёрский проект между MESR и OpenAlex, являющийся частью Второго национального плана открытой науки 2021 г., цель которого заключается в обеспечении гарантий, что данные, полученные в ходе французских государственных исследований, будут постепенно структурироваться в соответствии с принципами FAIR (доступность, функциональность, совместимость, возможность повторного использования), надёжно храниться и, по возможности, находиться в открытом доступе. Франция демонстрирует важность построения глобального ландшафта открытой науки. Следуя этой позиции, Парижский университет (Сорбонна) с 2024 г. отказался от подписки на Web of Science компании Clarivate в пользу OpenAlex¹⁰. Университет Сорбонны решил перенаправить свои усилия на изучение открытых и бесплатных инструментов, развитие международной альтернативы ресурсов, в частности, с помощью OpenAlex.

С 2023 г. появился Открытый выпуск Лейденского рейтинга университетов¹¹, который строится на библиографических данных из OpenAlex и предлагает полностью прозрачную информацию о научной деятельности более 1 400 крупнейших университетов мира. В рейтинге также используются данные из реестра организаций, созданного и поддерживаемого CWTS, который частично построен на данных из Реестра исследовательских организаций (ROR, <https://ror.org/>). Leiden Ranking Open Edition учитывает только часть публикаций в OpenAlex, которые определяются как основные публикации, т. е. публикации в международных научных журналах в областях, которые подходят для анализа цитирования и соответствуют следующим критериям: определённый тип документа (статьи или главы книги); указание авторов, аффилированных лиц и ссылок; публикация на английском языке; публикация не была отозвана и издана в профильном журнале.

⁸ Le CNRS se désabonne de la base de publications Scopus // Le CNRS : [сайт]. 2024. 11 janvier. URL: <https://cnrs.fr/fr/actualite/le-cnrs-se-desabonne-de-la-base-de-publications-scopus> (дата обращения: 31.07.2024).

⁹ French Ministry of Higher Education and Research partners with OpenAlex to develop a fully open bibliographic tool // Ministère de l'enseignement supérieur et de la recherche : [сайт]. 2024. February 15. URL: <https://ouvrirelascience.fr/french-ministry-of-higher-education-and-research-partners-with-openalex-to-develop-a-fully-open-bibliographic-tool> (дата обращения: 31.07.2024).

¹⁰ Sorbonne University unsubscribes from the Web of Science // Sorbonne Université : [сайт]. 2023. December 8. URL: <https://sorbonne-universite.fr/en/news/sorbonne-university-unsubscribes-web-science> (дата обращения: 31.07.2024).

¹¹ CWTS Leiden Ranking Open Edition : [сайт]. URL: <https://open.leidenranking.com> (дата обращения: 31.07.2024).

В ближайшей перспективе существенное влияние на развитие OpenAlex может оказать полученный в 2024 г. грант в размере 7,5 миллионов долларов от благотворительного фонда Arcadia (<https://arcadiafund.org.uk>), нацеленный на то, чтобы превратить OpenAlex в главный граф научных знаний для исследователей и организаций по всему миру¹².

ОБ OPENALEX

OpenAlex – относительно новый информационный ресурс, интерфейс которого лишь с 2023 г. стал удобен и понятен для пользователя, не требует дополнительных технических навыков для работы (ранее необходимы были навыки программирования на Python и R). OpenAlex активно развивается и уже стал популярным во всём мире и востребованным в условиях рестрикций к зарубежным лицензионным ресурсам в качестве открытого источника информации для российских пользователей наряду с Lens, Dimensions, Semantic Scholar, Internet Archive Scholar, BASE и др.

OpenAlex создан OurResearch – некоммерческой организацией, стремящейся сделать исследования открытыми и измерить воздействие научных документов в глобальной информационной системе. В 2024 г. в OpenAlex индексируется более 250 млн научных работ (журнальные статьи, материалы конференций, книги и главы книг, наборы данных и диссертации) из 250 тысяч источников. Первоначальным источником базы метаданных данных послужил проект каталогизации всей научной литературы в Интернете Microsoft Academic Graph (MAG), который прекратил своё существование в 2021 г., но стал ключевым источником данных о работах за предыдущие годы. Впоследствии актуальными крупнейшими ресурсами пополнения OpenAlex стали Crossref, arXiv, PubMed, ORCID, ROR, DOAJ, Unpaywall, Internet Archive и др.

Такие платформы, как Google Scholar и OpenAlex, сыграли важную роль в попытках оценить видимость и измерить влияние научных документов, но точное измерение зависит от уникальной идентификации всех участников научной экосистемы [7]. Всем объектам присваивается постоянный идентификатор OpenAlex, который действует как первичный ключ в наборе данных. Этот идентификатор выражается в виде URL-адреса, который может быть преобразован либо в удобочитаемое (веб-страница), либо в машиночитаемое (объект JSON) представление. Там, где это возможно, объектам также присваиваются идентификаторы из внешних систем, чтобы повысить совместимость. При отборе используется метод сопоставления метаданных с помощью идентификаторов (например, DOI – для публикаций, ROR-ID – для учреждений, ORCID – для авторов). При этом OpenAlex стремится связывать авторов работ с учреждениями, дополняет записи тематическими рубриками, данными о количестве цитирований, статусом открытого доступа, типом издания и др.

¹² OurResearch receives \$7.5M grant from Arcadia to establish OpenAlex, a milestone development for Open Science // OurResearch blog : [сайт]. 2024. March 13. URL: <https://blog.ourresearch.org/ourresearch-receives-7-5m-grant-from-arcadia-to-establish-openalex-a-milestone-development-for-open-science> (дата обращения: 31.07.2024).

Для создания графа научных знаний, связывающего публикации, их авторов, аффилированные организации и финансирование с обогащёнными метаданными используются передовые технологии (облачные вычисления и машинное обучение).

Экспорт результатов поиска осуществляется бесплатно по лицензии ССО, поэтому ими можно делиться и использовать для решения поставленных задач. В OpenAlex добавили возможность создания сложных поисковых запросов с логическими операторами AND (И), OR (ИЛИ), NOT (НЕТ), а также с учётом иных правил составления поисковых запросов (скобок и кавычек). Каждый тип объекта (произведения, авторы, учреждения, источники, издатели, концепции) имеет свою собственную конечную точку, к которой можно запрашивать различные фильтры для указания желаемого подмножества. Ещё одна особенность базы – привязка понятий (concepts) к объектам «Викиданных» (Wikidata) и статьям в «Википедии». Дерево понятий содержит около 65 тысяч понятий, которые присваиваются на основе их названий и аннотаций с использованием автоматизированного классификатора [8].

Наборы данных – важнейший источник информации. Уже во многих журналах появилась практика публикации не только текстов статей, но и ссылок на первичные данные, с определением вида публикации как data paper, data report, data article и data note. В OpenAlex же хранится информация о более чем 7,5 миллионах датасетов, из которых более 4,7 млн – в открытом доступе. Среди приоритетных тем наборов данных были определены следующие: процессы кристаллизации и контроль (921 244), геометрическая обработка изображений дистанционного зондирования Земли (289 490), порошковый дифракционный анализ (158 007) и др. Для России обнаружено 7 609 наборов данных по таким темам, как: применение частично упорядоченных множеств в химических исследованиях (304), органическая химия и катализ (147), инновации в химическом образовании и лабораторных методах (104), термофорез и термодиффузионные исследования (96), разнообразие и сохранение сосудистых растений Центральной Европы (94); и др.

Всем публикациям, индексируемым в OpenAlex, присваивается тема (topic), автоматически определяемая на основе названия публикации, аннотации, названия журнала и данных о цитирующих статьях. Всего выделено четыре домена со следующим распределением по количеству публикаций: физические науки (75 960 000), социальные науки (65 900 000), медицина (44 680 000), естественные науки (28 090 000), которые разделяются на предметные области и далее – на подобласти и более узкие темы. Распределение количества статей по темам визуализировано на иерархической диаграмме¹³.

OpenAlex обладает двумя основными преимуществами: лёгкий доступ к данным и широкий охват данных, которые широко используются в количественных научных исследованиях [9]. OpenAlex более доступен, чем базы данных на основе подписки, такие как Web of Science (WoS) и Scopus. Показано, что его охват шире, чем у некоторых традиционных библиометрических баз данных, таких как WoS [10]. В исследованиях делается вывод, что OpenAlex

¹³ Новый рубрикатор OpenAlex Topics: структурная диаграмма // RPubs : [сайт]. URL: https://rpubs.com/HSE_SC/oa-topics (дата обращения: 31.07.2024).

является расширенной версией Scopus и может быть надёжной альтернативой для некоторых видов анализа [6].

Всё большее число учёных использует OpenAlex в качестве источника данных и исследований [8; 11; 12]. Проведённый анализ [6] позволил определить, что OpenAlex превосходит Scopus по охвату, полезен для некоторых форм библиометрического анализа, в частности на уровне страны, представляет сопоставимые с данными, полученными с помощью Scopus. OpenAlex фиксирует больше идентификаторов ORCID, меньше рефератов и одинаковое количество информации открытого доступа на статью по сравнению с WoS и Scopus [5]. В условиях растущей тенденции к открытой науке появление этой платформы имеет решающее значение для содействия академическому общению и преодоления цифровых, технологических и интеллектуальных разрывов, существующих между странами и регионами или внутри них [1].

Платформа развивается, исследователи отмечают не только достоинства, но и проблемы, связанные, например, с неправильной классификацией статей [13], отсутствием учреждений или неполной институциональной информацией в метаданных журнальных статей в OpenAlex [9], отставанием по количеству проиндексированных ссылок (и, как следствие, цитирований) от Scopus [6]. Несмотря на имеющиеся проблемы, предлагаются наборы правил, позволяющие получать и использовать информацию из OpenAlex максимально точно и эффективно.

Рассмотрим возможности OpenAlex для поиска и оценки результатов научных исследований российских учёных.

МЕТОДИКА ИССЛЕДОВАНИЯ

В этом исследовании предлагается методология отслеживания публикаций открытого доступа российских учёных с использованием OpenAlex. В целях изучения эффективности использования OpenAlex нами был проанализирован общий массив представленных на платформе работ. Далее для получения данных об отражении работ российских исследователей на 01.07.2024 г. был сформулирован запрос по стране «Российская Федерация», с помощью дополнительных фильтров была собрана информация по годам публикаций, видам и наиболее актуальной тематике изданий, издателям, наличию наборов данных, наиболее цитируемым статьям (> 50) и их распределению по тематике и другим параметрам.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Данные с платформы OpenAlex Explorer (<https://alpha.openalex.org/>) были получены на 01.07.2024 г. с использованием различных фильтров. Для страны «Российская Федерация» был применён географический фильтр и проведён дополнительный поиск данных, касающихся российских учреждений.

Далее для получения более полного представления и достижения поставленных целей поисковый запрос уточнялся путём добавления других поисковых фильтров: «Вид документа», «Учреждение», «Автор», «Тема», «Открытый доступ», «Год», «Язык», «Издательство», «Количество цитирований».

Общий массив публикаций на 01.07.2024 г., отражённых в OpenAlex, составил 257 013 479; в нём преобладали статьи – 199 700 957, далее следовали главы из книг – 21 965 093, наборы данных – 7 528 992, диссертации – 6 057 627, препринты – 5 761 402 и др. Доля публикаций российских учёных составила 0,9% (2 398 923), среди которых больше всего статей – 2 210 172, глав из книг – 123 411, препринтов – 31 962 и наборов данных – 7 524. Отметим значительное увеличение числа публикаций в последние годы (рис. 1). В открытом доступе (ОД) доля публикаций составила почти 55% (849 032). Публикации на английском языке преобладают и составляют 94,3% (2 263 000), далее следует русский (3,7%) и немецкий языки (0,24%), 90 190 и 5 800 соответственно.

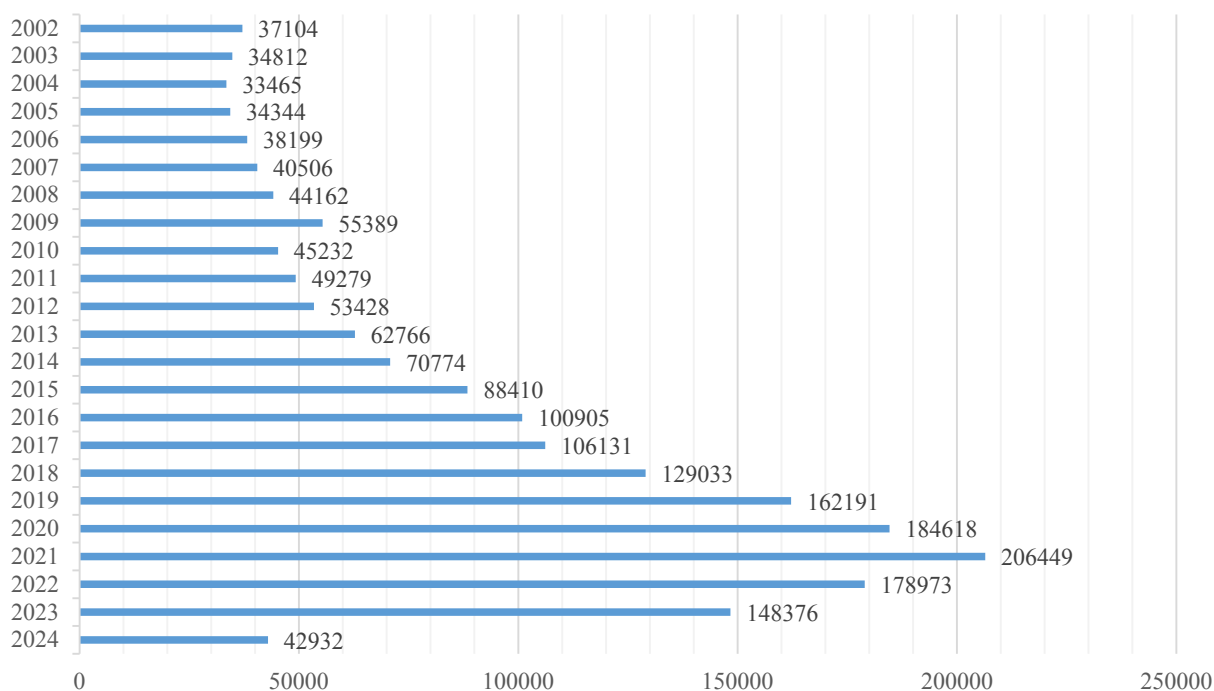


Рис. 1. Отражение результатов исследований российских учёных в OpenAlex по годам (2002–01.07.2024)

Fig. 1. Reflection of results of studies conducted by Russian researchers in OpenAlex by year (2002–01.07.2024)

Небольшое количество российских публикаций в OpenAlex (рис. 2) может быть обусловлено рядом факторов, например, отсутствием или неточностью представления данных об учреждении, что характерно не только для России, но и других стран, а также для иных, уже зарекомендовавших себя ресурсов, таких как WoS [10].

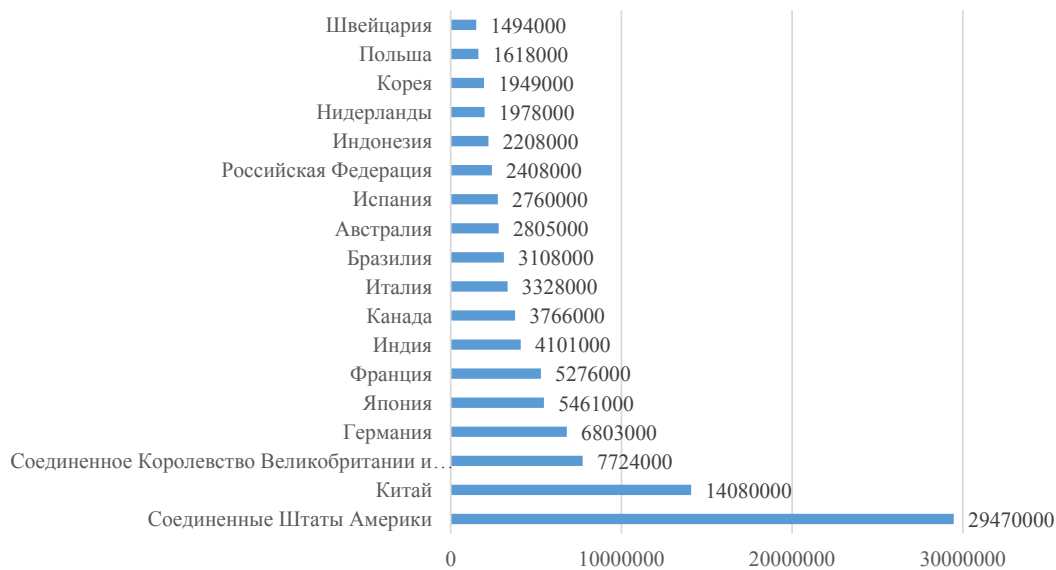


Рис. 2. Количество публикаций в OpenAlex по фильтру «Страна» (01.07.2024). Фрагмент
Fig. 2. Number of publications in OpenAlex by the filter “Country” (01.07.2024). Fragment

Для проведения более детального анализа были выбраны следующие параметры: учреждения, издательства, данные, тематика, авторы и цитируемость.

Учреждения. В OpenAlex индексируется более 100 тыс. учреждений, о принадлежности к которым заявляют авторы. В целях связывания учреждений с публикациями в системе анализируются все аффилиации, указанные каждым автором. Данные получают как из структурированных источников (например, PubMed), так и из неструктурированных (веб-страницы издателей). Около 94% учреждений в OpenAlex имеют уникальный идентификатор, полученный с помощью глобального реестра открытых постоянных идентификаторов для научных организаций (Research Organization Registry, <https://ror.org>), включающий данные более 110 000 учреждений. Например, ROR ID Института проблем технологии микроэлектроники и особочистых материалов РАН – <https://ror.org/034fqtg16>, а Физико-технического института Уральского отделения РАН – <https://ror.org/02jj4k517>. Данные реестра доступны в открытом доступе через интерфейс поиска. Информация извлекается, нормализуется и связывается с идентификаторами ROR с помощью алгоритмов машинного обучения. Вместе с тем результаты исследований показывают, что проблема отсутствия в публикациях учреждений встречается более чем в 60% журнальных статей в OpenAlex [9].

Поиск информации для анализа и определения трендов по учреждениям не всегда удобен. Так, по географическому параметру «Российская Федерация» в ТОП-200 учреждений входит категория «Российская академия наук» (949 112 публикаций, из них 273 500 в открытом доступе), где собраны различные институты РАН, например, Институт проблем технологии микроэлектроники и особочистых материалов, Санкт-Петербургское отделение Математического института им. В. А. Стеклова и др. На втором месте определена категория «Отдел физических наук», куда вошли 36 учреждений РАН с 235 417 работами (65 910 в ОД) учёных из Института кристаллографии им. А. В. Шубникова, Института ядерной физики им. Г. И. Будкера, Научного центра волоконной оптики, Института физики высоких давлений им. Л. Ф. Верещагина, Института астрономии и др.

На третьем месте представлена «Кафедра химии и материаловедения», в которой объединены 40 различных институтов РАН по общему направлению, в частности, Институт элементоорганических соединений им. А. Н. Несмеянова, Институт органической и физической химии им. А. Е. Арбузова, Иркутский институт химии им. А. Е. Фаворского, Институт нефтехимического синтеза им. А. В. Топчиева, Институт металлургии и материаловедения им. А. А. Байкова, Институт физической химии и электрохимии им. А. Н. Фрумкина и др., – с общим количеством публикаций 188 500 (27 180 в ОД).

В первой двадцатке учреждений значатся также отдельные университеты: Московский государственный университет им. М. В. Ломоносова (160 549), Санкт-Петербургский государственный университет (62 791), Национальный исследовательский университет «Высшая школа экономики» (47 558), Томский государственный университет (40 880), Новосибирский государственный университет (35 012), Казанский федеральный университет (30 915), Уральский федеральный университет (29 696). Эти университеты входят также в рейтинг российских университетов по доле публикаций в открытом доступе (по данным CWTS Leiden Ranking, 2023) за 2018–2021 гг.¹⁴

Издательства. Появление инициатив открытого доступа способствует преобразованию традиционной среды научного общения и приводит к увеличению количества журналов открытого доступа и исследовательских статей [1], в том числе издаваемых коммерческими организациями. В OpenAlex отражено большое количество публикаций таких известных издательств, как Elsevier, Springer Nature, Wiley и др. (рис. 3).

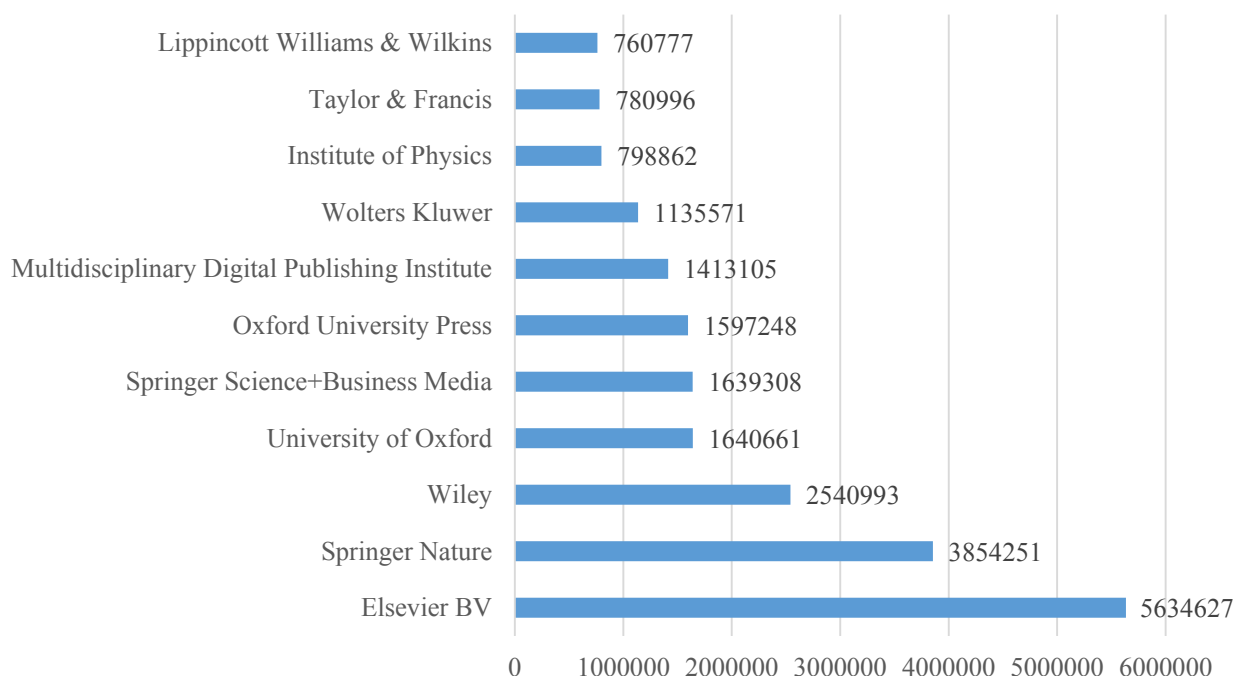


Рис. 3. Топ-10 издателей с наибольшим количеством статей в открытом доступе в OpenAlex
Fig. 3. Top 10 publishers with the largest number of open access articles in OpenAlex

¹⁴ CWTS Leiden Ranking : [сайт]. URL: <https://leidenranking.com> (дата обращения: 31.07.2024).

Больше всего статей российских учёных, нашедших отражение в OpenAlex, опубликовано в журналах издательской компании Springer Nature, группы издательств Pleiades Publishing, одного из крупнейших научных издательских домов мира Elsevier B.V., специализирующейся на выпуске академических журналов и книг по естественнонаучным направлениям международной издательской компании Springer Science+Business Media, британского научного общества Institute of Physics, издательской компании британского Института физики (IOP Publishing), издательства Wiley и др. (рис. 4).

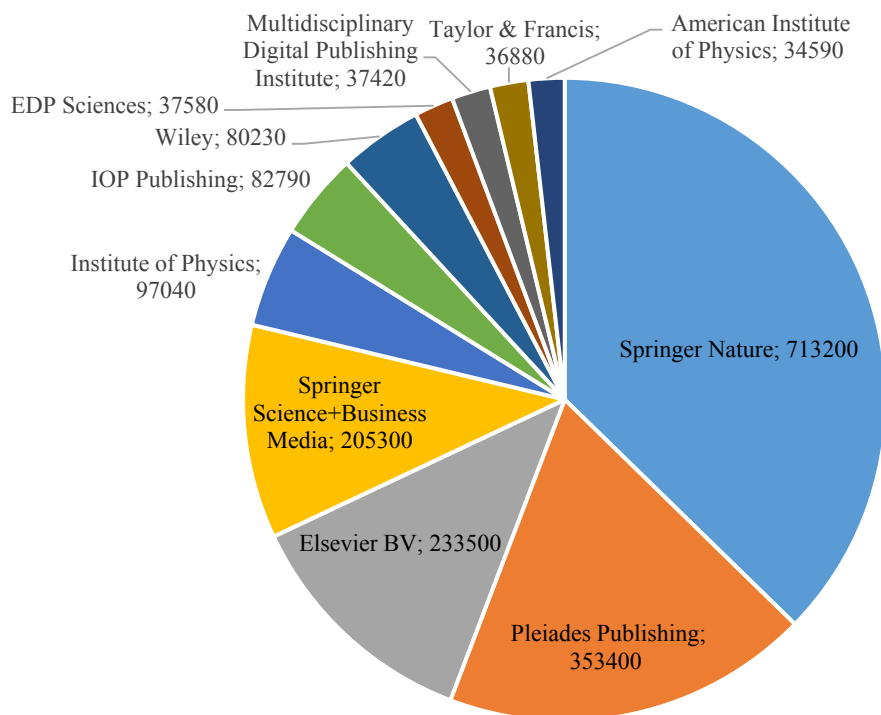


Рис. 4. Издательства, в которых опубликовано наибольшее количество статей российских учёных, отражённых в OpenAlex

Fig. 4. Publishers that released the largest number of Russian researchers' articles indexed in OpenAlex

Данные. Большое значение для развития открытой науки придаётся исследовательским данным, которые часто лежат в основе публикаций и могут быть доступны для проверки представленных результатов и облегчения повторного использования. Данные, отражённые в ресурсах открытого доступа, повышают прозрачность, экономят время других исследователей, занятых изучением той или иной проблемы, способствуют снижению риска потери данных и улучшают цитирование. В OpenAlex включены исследовательские данные по разным дисциплинам и научным отраслям, к примеру, наиболее широко представлены такие темы, как: «Процессы кристаллизации и контроль», «Геометрическая обработка изображений дистанционного зондирования Земли» и др. (табл. 1). Среди российских исследователей практика открытия данных пока менее распространена. Вместе с тем выделяются направления исследований, в которых есть ссылки на первичные данные исследований (табл. 2).

Таблица 1

Наиболее распространённая тематика наборов данных в OpenAlex (01.07.2024 г.)

Table 1

The most popular dataset topics in OpenAlex (01.07.2024)

Тематика наборов данных	Количество
Процессы кристаллизации и контроль	921 244
Геометрическая обработка изображений дистанционного зондирования Земли	289 490
Порошковый дифракционный анализ	158 007
Разнообразие и сохранение сосудистых растений Центральной Европы	24 018
Твёрдые кислоты в протонной проводимости и сегнетоэлектричестве	23 447
Разнообразие и эволюция грибковых патогенов	23 195
Позитронно-эмиссионная томография в онкологии	21 680
Разнообразие и филогения дождевых червей	21 184
Биоразнообразие сосудистых растений Кореи	14 546
Сосудистая флора Средиземноморья Европы и Северной Африки	13 363
Биологические эффекты электромагнитных полей	12 703
Разнообразие и применение видов <i>Superus</i>	12 545
Радиомика в анализе медицинских изображений	10 222

Таблица 2

Наиболее распространённая тематика наборов данных российских учёных в OpenAlex (01.07.2024 г.)

Table 2

The most popular dataset topics of Russian researchers in OpenAlex (01.07.2024)

Тематика наборов данных российских исследователей	Количество
Применение частично упорядоченных множеств в химических исследованиях	304
Органическая химия и катализ	147
Инновации в химическом образовании и лабораторных методах	104
Термофорез и термодиффузионные исследования	96
Разнообразие и сохранение сосудистых растений Центральной Европы	94
Каталитическое дегидрирование легких алканов	89
Геологическая эволюция Арктического региона	57
Разнообразие и эволюция мохообразных	52

Тематика. Количество статей по тематическим направлениям в журналах открытого и закрытого доступа различается. В открытом доступе преобладают публикации российских учёных по общественным и гуманитарным наукам, в то время как естественные и технические науки остаются в журналах закрытого доступа (рис. 5).



Рис. 5. Топ тематических направлений публикаций российских учёных в OpenAlex
Fig. 5. Top topics of publications by Russian researchers in OpenAlex

Авторы. OpenAlex индексирует более 257 млн авторов, причём ежедневно в связи с развитием ресурса добавляются тысячи учёных. OpenAlex использует, по возможности, ORCID, помогающий алгоритмически идентифицировать авторов. Система также анализирует записи публикаций авторов и историю цитирования для устранения неопределённости.

Самыми продуктивными российскими авторами по количеству публикаций, представленных в OpenAlex, стали С. И. Сухоручкин (9 922 глав книг, 44 статьи) и З. Н. Сороко (9 927 глав из книг, 14 статей). Большое количество публикаций обусловлено наличием совместных справочных материалов названных авторов, посвящённых различным химическим элементам, опубликованных как отдельные главы в электронных книгах закрытого доступа издательства Springer. На платформе Springer eBooks значится 2 519 000 аналогичных публикаций с общим количеством цитирований 11 720 000. Однако большое их количество не даёт высоких показателей цитируемости. Максимальный показатель – 12 для работы С. И. Сухоручкина и 27 для статьи З. Н. Сороко.

Цитируемость. Анализ цитирований позволяет оценить развитие научного направления, изучить взаимосвязи и влияние публикаций, авторов, учреждений, стран и журналов. Помимо уже зарекомендовавших себя источников библиометрического анализа, таких как Scopus и Web of Science, часто стал применяться OpenAlex, который даёт возможность проводить исследования цитируемости различных документов, организаций (количество цитирований работ организации с разбивкой по годам), издательств, тематик и пр., в т. ч. с использованием различных идентификаторов и полей, сортировкой результатов выдачи по количеству цитирований. К примеру, для анализа цитируемости публикаций Института проблем технологии микроэлектроники и особочистых материалов РАН можно составить более точное поисковое предписание с использованием фильтров «Учреждение» (<https://ror.org/034fqtg16>) и «Количество цитирований» (значение > 10). Возможны иные варианты изучения цитируемости в зависимости от цели исследования.

Поиск информации по запросу с использованием фильтров «Страна» («Российская Федерация») и «Количество цитирований» (значение > 50) дал следующие результаты: более 50 цитирований получили 64 522 статьи, 1 063 книги, 720 глав из книг, 292 препринта, 127 наборов данных других работ российских исследователей. В этот список попали 67 558 публикаций. Далее с помощью сортировки выданных результатов по теме определены наиболее цитируемые работы, которые посвящены изучению физики элементарных частиц и экспериментам на коллайдере высоких энергий, тектонической и геохронологической эволюции орогенов, формированию галактик, эволюции во Вселенной и др. (рис. 6).

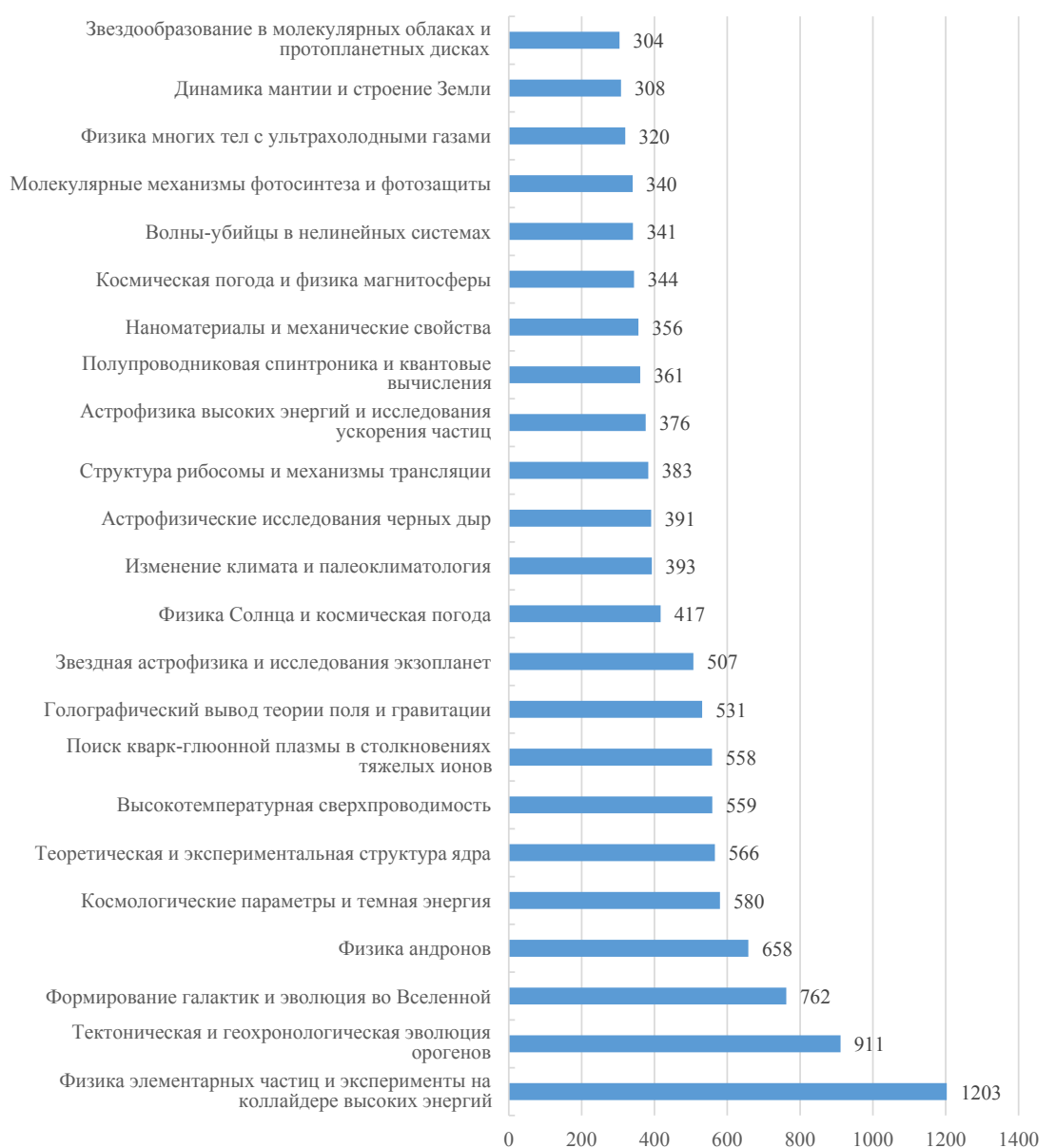


Рис. 6. Тематика наиболее цитируемых публикаций российских учёных в OpenAlex
Fig. 6. Topics of the most cited publications of Russian researchers in OpenAlex

Сортировка по количеству цитирований позволила получить данные о самых цитируемых работах с участием российских авторов и их тематике (химия материалов и молекулярная биология):

- Novoselov K. S. [et al.] Electric field effect in atomically thin carbon films. *Science*. 2004;306(5696):666–669. DOI 10.1126/science.1102896. Цитирований: 57 630 (статья в открытом доступе).
- Bankevich A. [et al.] SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 2012;19(5):455–477. DOI 10.1089/cmb.2012.0021. Цитирований: 19 550.
- Novoselov K. [et al.] Two-dimensional gas of massless Dirac fermions in graphene. *Nature*. 2005;438(7065):197–200. DOI 10.1038/nature04233. Цитирований: 19 430.

Для примера, наивысший показатель цитируемости имеет статья 1951 г., которая процитирована 311 100 раз. Она была подготовлена несколькими исследователями из Университета Вашингтона в Сент-Луисе (США) и опубликована в издании *Journal of Biological Chemistry*: Lowry O. H., Rosebrough N. J., Farr A. L., Randall R. J. Protein measurement with the Folin phenol reagent. *J Biol Chem.* 1951Nov;193(1):265–275. PMID 14907713.

Данные о цитировании не в полной мере отражают картину использования для каждой публикации, т. к. OpenAlex использует только те данные, которые взяты из открытых источников. Инициатива по открытому цитированию, предложенная независимой некоммерческой организацией OpenCitations (<https://opencitations.net/>), занимающейся публикацией открытых библиографических и цитируемых данных с использованием технологий Semantic Web (Linked Data) и пропагандой открытых цитирований, играет важную роль в том, чтобы издатели делали свою информацию о цитировании открытой, но, несмотря на инициативы, данные не являются полными на 100% в OpenAlex [14].

Таким образом, можно сделать вывод, что, несмотря на развитие принципов открытости в России, результаты исследований российских учёных пока недостаточно представлены в OpenAlex. Решению данной проблемы могут способствовать дальнейшее развитие систем идентификации для различных видов научной продукции и авторов (DOI, ORCID), индексирование российских ресурсов в каталогах (например, в DOAJ), реестрах и ресурсах открытого доступа (re3Data, OpenDOAR и др.¹⁵), повышение культуры открытых исследований, что позволит сделать их результаты доступными для обнаружения и последующего цитирования.

ЗАКЛЮЧЕНИЕ

OpenAlex – сравнительно недавняя инициатива, которая быстро развивается и завоёвывает признание по качеству и полноте предлагаемой информации благодаря открытости для исследовательского сообщества, повышению прозрачности оценки исследований, а также функциональным характеристикам и поисковым возможностям. Развитие сотрудничества с отдельными крупными издательствами, которые предоставляют данные своих журналов, сбор данных из выбранных доступных источников в Интернете в сочетании с разработанными алгоритмами, получение информации с других платформ, включая Crossref, PubMed и др., а также анализ каждой организации, указанной каждым автором в своих публикациях, позволяют сделать вывод о возможностях использования OpenAlex в качестве перспективного информационного ресурса для поиска разнородной информации и альтернативного инструмента библиометрического анализа, что особенно актуально для российских исследователей в условиях рестрикций. Однако для более эффективного применения OpenAlex и иных ресурсов открытого доступа в нашей стране требуется большее проникновение российской науки в информационную инфраструктуру открытой науки.

¹⁵ См.: Информационные ресурсы // Библиотека для открытой науки : [сайт]. URL: <https://lib-os.ru/issledovatelyam/resursy/> (дата обращения: 31.07.2024).

СПИСОК ИСТОЧНИКОВ

1. *Redkina N. S.* Impact of Open Science infrastructure on the development of the world information resources market // *Scientific and Technical Information Processing*. 2024. Vol. 51, № 2. P. 161–172. DOI 10.3103/S0147688224700096.
2. Literature reviews as independent studies: Guidelines for academic practice / *S. Kraus, M. Breier, W. M. Lim [et al.]* // *Review of Managerial Science*. 2022. Vol. 16, № 8. P. 2577–2595. DOI 10.1007/s11846-022-00588-8.
3. *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. DOI 10.19181/smtp.2023.5.2.4. EDN CXJUNG.
4. Recommendations for accelerating open preprint peer review to improve the culture of science / *M. Avissar-Whiting, F. Belliard, S. M. Bertozzi [et al.]* // *PLoS Biology*. 2024. Vol. 22, № 2. Article e3002502. DOI 10.1371/journal.pbio.3002502.
5. Reference coverage analysis of OpenAlex compared to Web of Science and Scopus / *J. Culbert, A. Hobert, N. Jahn [et al.]* // *arXiv : [сайт]*. 2024. January 29. URL: <https://arxiv.org/abs/2401.16359> (дата обращения: 31.07.2024). DOI 10.48550/arXiv.2401.16359.
6. An analysis of the suitability of OpenAlex for bibliometric analyses / *J. P. Alperin, J. Portenoy, K. Demes [et al.]* // *arXiv : [сайт]*. 2024. April 26. URL: <https://arxiv.org/abs/2404.17663> (дата обращения: 31.07.2024). DOI 10.48550/arXiv.2404.17663.
7. *Murillo-Gonzalez D., López S.* Persistent identifiers in the traceability of scientific research using OpenAlex // *2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)*, Cali, Colombia. Piscataway, NJ : Institute of Electrical and Electronics Engineers, 2023. P. 1–6. DOI 10.1109/AmITIC60194.2023.10366365.
8. *Priem J., Piwowar H., Orr R.* OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts // *26th International Conference on Science and Technology Indicators (STI 2022)*, September 7–9, 2022, Granada, Spain. Ed. by N. Robinson-Garcia, D. Torres-Salinas, W. Arroyo-Machado. Granada : University of Granada, 2023. Article sti22193. DOI 10.5281/zenodo.6936226.
9. Missing institutions in OpenAlex: Possible reasons, implications, and solutions / *L. Zhang, Z. Cao, Y. Shang [et al.]* // *Scientometrics*. 2024. February 5. DOI 10.1007/s11192-023-04923-y.
10. The concordance of field-normalized scores based on Web of Science and Microsoft Academic data: A case study in computer sciences / *T. Scheidsteger, R. Haunschild, S. Hug [et al.]* // *23rd International Conference on Science and Technology Indicators (STI 2018)*. September 12–14, 2018, Leiden, the Netherlands. Ed. by R. Costas, T. Franssen, A. Yegros-Yegros. Leiden : Centre for Science and Technology Studies (CWTS), Leiden University, 2018. P. 19–26. DOI 10.48550/arXiv.1802.10141.
11. *Haunschild R., Bornmann L.* Usage of OpenAlex for creating meaningful global overlay maps of science on the individual and institutional levels // *arXiv : [сайт]*. 2024. April 3. URL: <https://arxiv.org/abs/2404.02732> (дата обращения: 31.07.2024). DOI 10.48550/arXiv.2404.02732.
12. *Akbaritabar A., Theile T., Zagheni E.* Global flows and rates of international migration of scholars : MPIDR Working Paper WP-2023-018. Rostock : Max Planck Institute for Demographic Research, 2023. DOI 10.4054/MPIDR-WP-2023-018.
13. *Hauschke C., Nazarovets S.* (Non-)retracted academic papers in OpenAlex // *arXiv : [сайт]*. 2024. March 20. URL: <https://arxiv.org/abs/2403.13339> (дата обращения: 31.07.2024). DOI 10.48550/arXiv.2403.13339.
14. *Peroni S., Shotton D.* OpenCitations, an infrastructure organization for open scholarship // *Quantitative Science Studies*. 2020. Vol. 1, № 1. P. 428–444. DOI 10.1162/qss_a_00023.

REFERENCES

1. Redkina N. S. Impact of Open Science infrastructure on the development of the world information resources market. *Scientific and Technical Information Processing*. 2024;51(2):161–172. DOI 10.3103/S0147688224700096.
2. Kraus S., Breier M., Lim W. M. [et al.] Literature reviews as independent studies: guidelines for academic practice. *Review of Managerial Science*. 2022;16(8):2577–2595. DOI 10.1007/s11846-022-00588-8.
3. Gureev V. N., Mazov N. A. Increased role of open bibliographic data in the context of restricted access to proprietary information systems. *Science Management: Theory and Practice*. 2023;5(2):49–76. (In Russ.). DOI 10.19181/smtpr.2023.5.2.4.
4. Avissar-Whiting M., Belliard F., Bertozzi S. M. [et al.] Recommendations for accelerating open preprint peer review to improve the culture of science. *PLoS Biology*. 2024;22(2):e3002502. DOI 10.1371/journal.pbio.3002502.
5. Culbert J., Hobert A., Jahn N. [et al.] Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *arXiv*. 2024. January 29. Available at: <https://arxiv.org/abs/2401.16359> (accessed: 31.07.2024). DOI 10.48550/arXiv.2401.16359.
6. Alperin J. P., Portenoy J., Demes K. [et al.] An analysis of the suitability of OpenAlex for bibliometric analyses. *arXiv*. 2024. April 26. Available at: <https://arxiv.org/abs/2404.17663> (accessed: 31.07.2024). DOI 10.48550/arXiv.2404.17663.
7. Murillo-Gonzalez D., López S. Persistent identifiers in the traceability of scientific research using OpenAlex. In: 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), Cali, Colombia. Piscataway, NJ : Institute of Electrical and Electronics Engineers; 2023. P. 1–6. DOI 10.1109/AmITIC60194.2023.10366365.
8. Priem J., Piwowar H., Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. In: Robinson-Garcia N., Torres-Salinas D., Arroyo-Machado W., eds. 26th International Conference on Science and Technology Indicators (STI 2022). September 7–9, 2022, Granada, Spain. Granada : University of Granada; 2023. Article sti22193. DOI 10.5281/zenodo.6936226.
9. Zhang L., Cao Z., Shang Y. [et al.] Missing institutions in OpenAlex: Possible reasons, implications, and solutions. *Scientometrics*. 2024. February 5. DOI 10.1007/s11192-023-04923-y.
10. 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: Costas R., Franssen T., Yegros-Yegros A., eds. 23rd International Conference on Science and Technology Indicators (STI 2018). September 12–14, 2018, Leiden, the Netherlands. Leiden : Centre for Science and Technology Studies (CWTS), Leiden University; 2018. P. 19–26. DOI 10.48550/arXiv.1802.10141.
11. Haunschild R., Bornmann L. Usage of OpenAlex for creating meaningful global overlay maps of science on the individual and institutional levels. *arXiv*. 2024. April 3. Available at: <https://arxiv.org/abs/2404.02732> (accessed: 31.07.2024). DOI 10.48550/arXiv.2404.02732.
12. Akbaritabar A., Theile T., Zagheni E. Global flows and rates of international migration of scholars : MPIDR Working Paper WP-2023-018. Rostock : Max Planck Institute for Demographic Research; 2023. DOI 10.4054/MPIDR-WP-2023-018.
13. Hauschke C., Nazarovets S. (Non-)retracted academic papers in OpenAlex. *arXiv*. 2024. March 20. Available at: <https://arxiv.org/abs/2403.13339> (accessed: 31.07.2024). DOI 10.48550/arXiv.2403.13339.
14. Peroni S., Shotton D. OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*. 2020;1(1):428–444. DOI 10.1162/qss_a_00023.

Поступила в редакцию / Received 02.08.2024.

Одобрена после рецензирования / Revised 15.08.2024.

Принята к публикации / Accepted 12.11.2024.

СВЕДЕНИЯ ОБ АВТОРЕ**Редькина Наталья Степановна** *redkina@spsl.nsc.ru*

Доктор педагогических наук, заведующий отделом, Государственная публичная научно-техническая библиотека СО РАН, Новосибирск, Россия

SPIN-код: 9887-6329

INFORMATION ABOUT THE AUTHOR**Natalya S. Redkina** *redkina@spsl.nsc.ru*

Doctor of Pedagogy, Department Head, State Public Scientific Technological Library, SB RAS, Novosibirsk, Russia

ORCID: 0000-0002-3486-9711

Web of Science ResearcherID: I-9602-2018

Scopus Author ID: ID: 57210434112