

РОССИЯ В МИРОВОМ МАССИВЕ НАУЧНЫХ ПУБЛИКАЦИЙ

© 2025 г. Ю.В. Мохначева^{а,*}

^аБиблиотека по естественным наукам РАН, Москва, Россия

*E-mail: mohnacheva@benran.ru

Поступила в редакцию 03.10.2024 г.
После доработки 29.10.2024 г.
Принята к публикации 02.12.2024 г.

Выбор наиболее релевантной платформы для проведения мониторинговых исследований российского массива публикаций на мировом фоне, включая анализ публикационных потоков по различным научным категориям, — актуальная задача. Объектами проведённого исследования служили российский и мировой массивы публикаций в базах данных OpenAlex и The Lens в сравнении с Web of Science Core Collection (WoS CC) за период 2014–2023 гг. с особым акцентом на период 2019–2023 гг. Обнаружено, что OpenAlex обладает рядом преимуществ перед The Lens, поскольку российские публикации значительно недопредставлены в The Lens. Сравнительный анализ динамики долей российских публикаций по предметным категориям в OpenAlex и The Lens на фоне WoS CC показал, что в течение 2019–2023 гг. разрыв между долями российских публикаций в мировом массиве в этих ресурсах относительно WoS CC по большинству пересекающихся предметных категорий сократился. Данные о долевого распределении российских публикаций в 2021–2023 гг. в OpenAlex коррелируют с показателями в WoS CC, что позволяет надеяться на высокую релевантность результатов поиска с помощью OpenAlex.

Ключевые слова: наука в России, вклад России в мировую науку, открытые информационно-библиографические базы данных, массивы публикаций, OpenAlex, The Lens, наукометрические исследования, предметные категории.

DOI: 10.31857/S0869587325010052, EDN: ANCSLQ

С момента введения беспрецедентных по масштабу антироссийских санкций со стороны США, Европейского Союза и других недружественных стран российским пользователям на неопределённый срок закрыт доступ к наиболее известным и авторитетным ресурсам поиска научно-технической и наукометрической информации — Web of Science (Clarivate, США) и Scopus (Elsevier, Нидерланды). Несмотря на некоторые недостатки этих баз данных [1–4], на их основе генерируются отчёты о пуб-

ликационной активности на всех уровнях: микро- (уровень отдельных персоналий и публикаций), мезо- (уровень организаций) и макро- (уровень государств). С уходом этих баз данных из российского информационного пространства с особой остротой встала задача поиска альтернативных путей получения сведений о публикационной активности, особенно на макроуровне.

Для осуществления успешной научной политики необходимы релевантные данные о состоянии и динамике российских публикационных массивов на общемировом фоне, включая сведения по отдельным научным направлениям, в том числе приоритетным. Авторитет Web of Science (WoS) и Scopus сформирован главным образом благодаря широкому функционалу и качественному контенту: один только факт включения журналов в эти базы данных подразумевает высокое качество источников, поскольку основан на строгих правилах отбора [1]. Тем не менее настало время для поиска альтернативы этим системам.

Хотя в распоряжении пользователей имеется достаточно широкий круг открытых ресурсов



МОХНАЧЕВА Юлия Валерьевна — кандидат педагогических наук, ведущий научный сотрудник БЕН РАН.

(OpenAlex, The Lens, Dimensions, Scilit, Semantic Scholar, РИНЦ, CoLab и др.), которые позволяют получать разноплановую информацию о публикациях, включая библиометрические и альтметрические данные [5, 6], далеко не все системы отвечают необходимым требованиям и располагают полноценным функционалом. Возможности таких ресурсов часто ограничены информацией о публикациях на уровне авторов и их публикаций, реже — на уровне организаций и совсем редко — на уровне государств.

В создании и развитии библиографических и наукометрических систем в России достигнут значительный прогресс. Успешно действуют и активно развиваются такие базы данных, как РИНЦ на платформе eLibrary.ru (ООО “Научная электронная библиотека”)¹, CoLab², КиберЛенинка³, портал “Белый список” журналов (РЦНИ)⁴, система ИСТИНА (МГУ им. М.В. Ломоносова)⁵ и др. Министр науки и высшего образования РФ В.Н. Фальков 27 сентября 2024 г. анонсировал создание единой наукометрической базы данных стран БРИКС [7]. Несмотря на широкий круг открытых российских и зарубежных библиографических ресурсов, лишь два из них — OpenAlex и The Lens — позволяют определять статус России в мировом научном пространстве.

Ресурсы открытого или частично открытого доступа обладают очевидными достоинствами и преимуществами — бесплатностью, универсальностью, политематичностью, широтой охвата мирового научного контента. Библиографические информационные ресурсы на основе поисковых систем и веб-сканирования, включая открытые, обеспечивают более полный охват документов, чем WoS и Scopus [8]. Тем не менее к таким системам предъявляют ряд вопросов: насколько высока авторитетность индексируемых источников; насколько корректны метаданные; насколько прозрачна политика и методика отбора и индексирования источников и т.д.? Несоответствия в метаданных публикаций в библиографических ресурсах негативно отражаются на релевантности и точности получаемых результатов [1]. Справедливости ради следует отметить, что “эталонные” WoS и Scopus также не лишены недостатков [9–11].

Сравнение различных баз данных между собой — самостоятельное направление исследований в информационно-библиотечной области. Ресурсы оценивают по различным критериям — от охвата индексируемых изданий и глубины ретроспективы наполнения баз данных до различий в методологии

типизации документов и предметной классификации, а также возможностей применения в информационно-библиографической деятельности [1, 2, 6, 9–15].

Авторы публикации [8], чтобы сопоставить охват различных научных баз данных, опираясь на принцип случайной выборки из Crossref, предположили, что использование стороннего третьего ресурса для сравнения охвата двух других уменьшит возможную предвзятость и позволит узнать, как критерии отбора и технические требования влияют на охват научной литературы. Исследователи пришли к выводу, что с позиции Crossref существуют значительные различия в охвате между научными базами данных, причём эти различия, как правило, можно объяснить разницей в методологии при создании баз данных, включая типизацию документов. В работе [2] сопоставлены пять многопрофильных библиографических ресурсов за период 2008–2017 гг.: Scopus, Web of Science, Dimensions, Crossref и Microsoft Academic. Каждая из систем сравнивалась со Scopus. Авторы делают акцент на проблемах, связанных с классификацией типов документов, и отмечают сильную зависимость Dimensions от данных из Crossref.

Ошибки в аффилиациях авторов — общая существенная проблема проприетарных ресурсов и систем открытого доступа [2]. Такие неточности напрямую влияют на статистику публикационной активности на мезо- и макроуровне; вероятно, они возникают на этапе индексирования публикаций информационными системами. Этой проблемой обеспокоены все без исключения производители библиографических баз данных, однако она до сих пор не решена.

В 2015–2021 гг. заслуженной популярностью пользовался открытый ресурс Microsoft Academic. Согласно выводам, представленным в статье [2], эта база охватывала и индексировала более широкий круг документов, чем другие источники данных. Превалирующая доля приходилась на научные работы, что было доказано результатами ручной выборки. К наиболее сильной стороне Microsoft Academic специалисты относили технологию сканирования веб-страниц с последующим применением искусственного интеллекта для актуализации базы данных исчерпывающими метаданными, что позволило повысить скорость индексирования новых документов [16]. Ещё одно достоинство Microsoft Academic — большое число отражённых документов не на английском языке. В процессе поиска новых способов расширения возможностей исследователей был реализован проект в двух технологических решениях: сначала считывались все проиндексированные Bing веб-страницы, отбиралась наиболее актуальная научная информация и генерировалась база данных под названием

¹ РИНЦ. <https://elibrary.ru/defaultx.asp>

² CoLab. <https://colab.ws/>

³ Научная электронная библиотека “КиберЛенинка”. <https://cyberleninka.ru/>

⁴ “Белый список” научных журналов РЦНИ. <https://journalrank.rcsi.science/ru/>

⁵ ИСТИНА. <https://istina.msu.ru/>

Microsoft Academic Graph (MAG)⁶; далее осуществлялось семантическое обоснование и вывод для обслуживания этих знаний посредством поиска на веб-сайте Microsoft Academic [17]. Решение разработчиков о прекращении поддержки Microsoft Academic в мае, а MAG в декабре 2021 г. вызвало серьёзную тревогу пользователей [18].

В 2022 г. в качестве альтернативы Microsoft Academic и MAG американская компания OurResearch⁷ запустила систему OpenAlex, названную так в честь древней Александрийской библиотеки. OpenAlex⁸ — это бесплатный и полностью открытый каталог научных метаданных с открытым исходным кодом, позволяющий проводить комплексный библиографический анализ. Как и Google Scholar, OpenAlex не использует экспертные оценки при отборе источников, предпочитая им широту охвата контента [19–21].

OpenAlex собирает и стандартизирует данные из многих источников, в первую очередь из MAG и Crossref⁹, а также данные ORCID¹⁰, ROR¹¹, DOAJ¹², Unpaywall¹³, Pubmed¹⁴, Pubmed Central¹⁵, The ISSN International Centre¹⁶, различных репозиториях. OpenAlex индексирует более 240 млн работ, ежедневно добавляется около 50 000 документов [21, 22]. В 2024 г. Лейденский университет (Нидерланды) для составления рейтинга научно-исследовательских учреждений использовал OpenAlex [23, 24]. OpenAlex позволяет выгружать необходимые данные как по API, так и путём обычной выгрузки данных со страницы результатов поиска в формате CSV. Важное достоинство системы — возможность получать сведения о публикационных массивах на всех уровнях: микро-, мезо- и макро-.

База данных The Lens¹⁷ создана в 1998 г. Это бесплатная платформа для поиска научной и патентной литературы, совместный проект некоммерческой организации Cambia и Технологического университета Квинсленда (Австралия) [25]. В ней реализована концепция MetaRecord (MeR), которая позволяет управлять проблемами, связанными с изменчивостью записей, источниками и контекстной релевантностью метаданных к исходной записи [26].

⁶ Microsoft Academic Graph (MAG). <https://www.microsoft.com/en-us/research/project/academic/>

⁷ Известна в научно-издательском сообществе базой данных и плагином Unpaywall.

⁸ OpenAlex. <https://openalex.org/works>

⁹ Crossref. <https://www.crossref.org/>

¹⁰ Open Researcher and Contributor ID (ORCID — от англ. “Открытый идентификатор исследователя и участника”). <https://orcid.org/>

¹¹ ROR (Research Organization Registry). <https://ror.org/>

¹² DOAJ (Directory of Open Access Journals). <https://doaj.org/>

¹³ Unpaywall. <https://unpaywall.org/>

¹⁴ Pubmed. <https://pubmed.ncbi.nlm.nih.gov/>

¹⁵ Pubmed Central. <https://www.ncbi.nlm.nih.gov/pmc/>

¹⁶ ISSN International Centre. <https://www.issn.org/>

¹⁷ The Lens. <https://www.lens.org/>

The Lens использует информацию сторонних систем (таких как PubMed и Crossref), объединяя их в одну, дедулицированную с унифицированным синтаксисом поиска базу данных. В отличие от конкурирующих баз данных, ресурс позволяет экспортировать данные в формате JSON с более подробной детализацией по сравнению с форматами RIS и CSV [27].

В настоящее время OpenAlex и The Lens, по сути, единственные открытые ресурсы, благодаря которым можно находить и выгружать данные не только об отдельных публикациях, авторах и научных организациях, но и о публикационных массивах на макроуровне.

Методология исследования. В качестве объекта исследования выступал российский массив публикаций в базах данных OpenAlex и The Lens в сравнении с WoS CC. Цель исследования — выбор оптимальной системы для мониторинга российского массива публикаций, а также определение доли российских публикаций в мировом научно-информационном поле по открытым базам данных OpenAlex и The Lens в сравнении с данными WoS CC за 2014–2023 гг. (акцент на 2019–2023 гг.).

В соответствии с утверждением, что обращение к стороннему третьему ресурсу для сравнения охвата двух других уменьшает возможные предвзятости [2], в качестве эталонных были приняты данные о динамике российского массива публикаций на мировом фоне в WoS CC. Использовался открытый модуль WoS CC, позволяющий проводить поиск¹⁸ без возможности выгрузки информации в файл. Данные собирались по каждому из трёх ресурсов WoS CC, OpenAlex и The Lens по миру в целом и по России в частности за каждый год периода 2014–2023 гг., в том числе по пересекающимся предметным категориям.

Для понимания динамических характеристик российского документопотока, учитывая активно сужающийся круг ресурсов как для авторов публикаций, так и для специалистов в области науки, метрии и научных администраторов, особенно важен временной интервал 2019–2023 гг. Поэтому исследование проведено с акцентом на этот период.

Определённые затруднения были связаны с тем, что базы данных используют собственные классификаторы, а значит, не все названия предметных категорий в одном ресурсе полностью идентичны категориям в другом. Поэтому здесь представлены результаты сравнения баз данных только в пересекающихся научных категориях (табл. 1–4).

¹⁸ Опция поиска по стране доступна в открытом модуле WoS CC через переход по активным ссылкам, приведённым в карточке публикации в WoS CC (при условии регистрации и авторизации в системе). Далее посредством активизации поискового поля возможен выход в расширенное меню поиска по всем полям.

Таблица 1. Пересекающиеся предметные категории WoS CC и OpenAlex, по которым доля российских публикаций на мировом фоне в WoS CC выше, чем в OpenAlex, 2019–2023 гг. (нисходящая сортировка по доле российских публикаций в WoS CC с минимальным порогом 1.5% мирового массива)

Предметная категория		Всего публикаций в мире за период		Всего российских публикаций за период		Доля российских публикаций, %		Преимущество доли российского массива в WoS CC над OpenAlex (величина отношения доли в WoS CC и в OpenAlex)
WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	
Physics, Nuclear	Nuclear and High Energy Physics	39571	161860	4784	9002	12.1	5.6	2.2
Paleontology	Paleontology	19367	53286	1767	1807	9.1	3.4	2.7
Chemistry, Inorganic & Nuclear	Inorganic Chemistry	75888	105910	5622	3848	7.4	3.6	2.1
Physics, Mathematical	Mathematical Physics	65239	89690	4773	3987	7.3	4.5	1.6
Astronomy & Astrophysics	Astronomy and Astrophysics	135094	318820	9596	12389	7.1	3.9	1.8
Spectroscopy	Spectroscopy	40176	89490	2571	3478	6.4	3.9	1.6
Chemistry, Organic	Organic Chemistry	100573	325480	5469	10948	5.4	3.4	1.6
Physics, Condensed Matter	Condensed Matter Physics	236135	79510	12816	3236	5.4	4.1	1.3
Oceanography	Oceanography	59583	161690	3197	6039	5.4	3.7	1.5
Mathematics, Applied	Applied Mathematics	227173	146230	10374	4905	4.6	3.4	1.4
History	History	272941	295320	11697	1844	4.3	0.6	7.2
Archaeology	Archeology	43272	249140	1782	3029	4.1	1.2	3.4
Soil Science	Soil Science	38871	140690	1597	4061	4.1	2.9	1.4
Language & Linguistics	Language and Linguistics	97338	283330	3945	5389	4.1	1.9	2.2
Zoology	Animal Science and Zoology	106204	85850	4111	1148	3.9	1.3	3
Ecology	Ecology	148089	360180	5331	7767	3.6	2.2	1.6
Economics	Economics and Econometrics	265931	808200	8589	11642	3.2	1.4	2.3
Instruments & Instrumentation	Instrumentation	182888	18893	5573	307	3.1	1.6	1.9

Окончание таблицы 1 на стр. 52.

Таблицы 1 (окончание)

Предметная категория		Всего публикаций в мире за период		Всего российских публикаций за период		Доля российских публикаций, %		Преимущество доли российского массива в WoS CC над OpenAlex (величина отношения доли в WoS CC и в OpenAlex)
WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	
Philosophy	Philosophy	142785	272840	4235	1998	3	0.7	4.3
Forestry	Forestry	47444	62620	1359	142	2.9	0.2	14.5
Sociology	Sociology and Political Science	103373	2020900	2790	25932	2.7	1.3	2.1
Education & Educational Research	Education	287957	1318600	7510	11164	2.6	0.9	2.9
History & Philosophy of Science	History and Philosophy of Science	44167	133520	1053	1016	2.4	0.8	3
Physiology	Physiology	82140	348440	1782	5233	2.2	1.5	1.5
Information Science & Library Science	Library and Information Sciences	72565	34203	1204	46	1.7	0.1	17
Anthropology	Anthropology	60022	233930	974	1969	1.6	0.8	2
Developmental Biology	Developmental Biology	32300	10103	502	120	1.6	1.2	1.3
Transplantation	Transplantation	73582	14214	1103	114	1.5	0.8	1.9

При сборе информации учитывались все типы документов, поскольку замечено, что от ресурса к ресурсу одни и те же работы могут относиться к разным типам документов [9–11]. В WoS CC поиск проводился одновременно по всем индексам по состоянию на июль–август 2024 г.: Science Citation Index Expanded; Social Sciences Citation Index; Arts & Humanities Citation Index; Emerging Sources Citation Index; Conference Proceedings Citation Index – Science; Conference Proceedings Citation Index – Social Science & Humanities.

Особенности представления информации в OpenAlex и The Lens. Для понимания широты охвата контента в OpenAlex и The Lens на фоне WoS CC, на рисунке 1

представлена динамика числа публикаций в мире за период 2014–2023 гг. (по состоянию на июль–август 2024 г.). Как видим, по числу публикаций открытые ресурсы OpenAlex и The Lens во много раз превышают статистику по WoS CC. При этом, за исключением 2023 г., динамика публикаций в OpenAlex и The Lens практически идентична, что наводит на мысль о возможной высокой степени схожести контента этих ресурсов. Однако данное предположение опровергают как результаты исследования динамики российского массива публикаций (рис. 2), так и анализ публикационных массивов по пересекающимся предметным категориям (табл. 3, 4).

Таблица 2. Пересекающиеся предметные категории WoS CC и OpenAlex, по которым доля российских публикаций на мировом фоне в OpenAlex выше, чем в WoS CC, 2019–2023 гг. (нисходящая сортировка по доле российских публикаций в OpenAlex с минимальным порогом 1.5% мирового массива)

Предметная категория		Всего публикаций в мире		Всего российских публикаций		Доля российских публикаций		Преимущество доли российского массива в OpenAlex (величина отношения доли в OpenAlex и в WoS CC)
OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	
Development	Development Studies	76750	33354	11739	1254	15.3	3.8	4
Anatomy	Anatomy & Morphology	21441	19248	1832	307	8.5	1.6	5.3
Agronomy and Crop Science	Agronomy	109730	102145	7289	1622	6.6	1.6	4.1
Cultural Studies	Cultural Studies	248900	52865	12721	110	5.1	0.2	25.5
Mechanical Engineering	Engineering, Mechanical	625200	237711	29622	7167	4.7	3	1.6
Acoustics and Ultrasonics	Acoustics	5340	52864	224	1191	4.2	2.3	1.8
Food Science	Food Science & Technology	302160	243197	11120	3170	3.7	1.3	2.8
Developmental and Educational Psychology	Psychology, Developmental	179690	59895	6498	119	3.6	0.2	18
Demography	Demography	306280	20364	10718	189	3.5	0.9	3.9
Analytical Chemistry	Chemistry, Analytical	68950	202464	2162	4248	3.1	2.1	1.5
Toxicology	Toxicology	15977	98343	352	755	2.2	0.8	2.8
Transportation	Transportation	81910	94936	1793	911	2.2	1	2.2
Pediatrics, Perinatology and Child Health	Pediatrics	317690	210115	6276	973	2	0.5	4
Ophthalmology	Ophthalmology	113080	121952	1995	478	1.8	0.4	4.5
Radiology, Nuclear Medicine and Imaging	Radiology, Nuclear Medicine & Medical Imaging	374070	246871	6430	1551	1.7	0.6	2.8
Urology	Urology & Nephrology	48136	147802	829	1476	1.7	1	1.7
Immunology	Immunology	238860	255004	4044	2615	1.7	1	1.7
Infectious Diseases	Infectious Diseases	337000	146706	4942	1363	1.5	0.9	1.7

Таблица 3. Пересекающиеся предметные категории OpenAlex и The Lens, по которым доля российских публикаций в OpenAlex выше, чем в The Lens, на мировом фоне, 2019–2023 гг. (нисходящая сортировка по доле российских публикаций в OpenAlex с минимальным порогом 1.5% мирового массива)

Предметная категория OpenAlex/The Lens	Всего публикаций в мире		Всего российских публикаций		Доля российских публикаций		Преимущество доли российского массива от мирового в OpenAlex над The Lens (величина отношения доли в OpenAlex и в TheLens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Fuel Technology	6338	135808	1510	3502	23.8	2.6	9.2
Nuclear Energy and Engineering	3843	58093	876	1651	22.8	2.8	8.1
General Materials Science	32629	753923	6509	12957	20	1.7	11.8
Geology	71610	108917	5670	2661	7.9	2.4	3.3
Mechanics of Materials	267620	367260	19429	8118	7.3	2.2	3.3
Agronomy and Crop Science	109730	193636	7289	667	6.6	0.3	22
Earth-Surface Processes	48576	88055	3202	2083	6.6	2.4	2.8
Ceramics and Composites	24267	123200	1517	2839	6.3	2.3	2.7
General Energy	16274	96058	991	1082	6.1	1.1	5.5
General Economics, Econometrics and Finance	168900	75674	9897	1052	5.9	1.4	4.2
Atomic and Molecular Physics, and Optics	297970	292242	16919	9567	5.7	3.3	1.7
Cultural Studies	248900	237453	12721	2712	5.1	1.1	4.6
Industrial and Manufacturing Engineering	205360	521695	10336	7384	5	1.4	3.6
Mechanical Engineering	625200	540833	29622	9574	4.7	1.8	2.6
Catalysis	41055	191677	1691	2883	4.1	1.5	2.7
Atmospheric Science	204490	83959	8361	1807	4.1	2.2	1.9
Spectroscopy	89490	156072	3478	3158	3.9	2	2
Physiology (Medicine)	18956	129407	730	946	3.9	0.7	5.6
Food Science	302160	259669	11120	875	3.7	0.3	12.3
Organic Chemistry	325480	356484	10948	6528	3.4	1.8	1.9
Control and Systems Engineering	363210	219988	11488	2698	3.2	1.2	2.7
Analytical Chemistry	68950	208691	2162	2212	3.1	1.1	2.8

Продолжение таблицы 3 на стр. 55.

Таблица 3 (продолжение)

Предметная категория OpenAlex/The Lens	Всего публикаций в мире		Всего российских публикаций		Доля российских публикаций		Преимущество доли российского массива от мирового в OpenAlex над The Lens (величина отношения доли в OpenAlex и в TheLens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Biophysics	48194	122030	1467	1867	3	1.5	2
Information Systems	809600	148238	24493	1199	3	0.8	3.8
Soil Science	140690	77988	4061	1056	2.9	1.4	2.1
Neurology (Medicine)	97190	203879	2724	801	2.8	0.4	7
Ecology, Evolution, Behavior and Systematics	236010	312256	6554	4615	2.8	1.5	1.9
Biomedical Engineering	694200	163838	17959	1631	2.6	1	2.6
Polymers and Plastics	116770	321306	3016	4177	2.6	1.3	2
Strategy and Management	400480	214984	10092	828	2.5	0.4	6.3
Management, Monitoring, Policy and Law,	305050	225423	7648	1636	2.5	0.7	3.6
Electrochemistry	16220	76489	403	907	2.5	1.2	2.1
Pharmaceutical Science	45256	240888	1123	1875	2.5	0.8	3.1
Electrical and Electronic Engineering	1165300	771632	28597	8295	2.5	1.1	2.3
General Social Sciences	72816	82197	1758	816	2.4	1	2.4
Law	404280	153243	9684	1216	2.4	0.8	3
Bioengineering	15572	138816	362	1133	2.3	0.8	2.9
Building and Construction	233360	186502	5405	1062	2.3	0.6	3.8
Computational Theory and Mathematics	194420	76846	4509	654	2.3	0.9	2.6
Ecology	360180	202313	7767	2312	2.2	1.1	2
Cardiology and Cardiovascular Medicine	389200	434927	7797	1751	2	0.4	5
Pediatrics, Perinatology and Child Health	317690	321291	6276	690	2	0.2	10
Cellular and Molecular Neuroscience	106820	106528	2108	501	2	0.5	4
Biochemistry	33775	616063	648	6331	1.9	1	1.9
Physical Therapy, Sports Therapy and Rehabilitation	94410	116473	1816	385	1.9	0.3	6.3

Окончание таблицы 3 на стр. 56.

Таблица 3 (окончание)

Предметная категория OpenAlex/The Lens	Всего публикаций в мире		Всего российских публикаций		Доля российских публикаций		Преимущество доли российского массива от мирового в OpenAlex над The Lens (величина отношения доли в OpenAlex и в TheLens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Civil and Structural Engineering	423680	238539	7950	898	1.9	0.4	4.8
Genetics	353910	309737	6601	2090	1.9	0.7	2.7
Internal Medicine	29023	116339	536	355	1.9	0.3	6.3
Microbiology	44600	151286	824	1129	1.9	0.8	2.4
Renewable Energy, Sustainability and the Environment	280460	319956	5191	1902	1.9	0.6	3.2
Ophthalmology	113080	112170	1995	392	1.8	0.4	4.5
Immunology and Allergy	35945	232002	625	1430	1.7	0.6	2.8
Gastroenterology	46920	183238	798	388	1.7	0.2	8.5
Immunology	238860	282911	4044	1766	1.7	0.6	2.8
Molecular Biology	1220200	501775	19438	3733	1.6	0.7	2.3
Biotechnology	58670	233457	929	1456	1.6	0.6	2.7
Environmental Engineering	180530	208640	2783	1946	1.5	0.9	1.7
Physiology	348440	156046	5233	1305	1.5	0.8	1.9
Computer Networks and Communications	359870	230365	5363	1018	1.5	0.4	3.8
Infectious Diseases	337000	257944	4942	1099	1.5	0.4	3.8

Из рисунка 2 видно, что лидерство по индексации российских публикаций принадлежит OpenAlex по каждому году периода 2014–2023 гг., причём после 2017 г. это особенно заметно. Снижение динамики в последние два года на контрасте с активным ускорением в 2017–2021 гг., характерным для всех трёх систем, по-видимому, связано с техническими задержками в индексировании документов. До 2019 г. WoS CC по числу российских документов опережал The Lens, но начиная с 2020 г. рост российского массива публикаций в WoS CC заметно замедлился. Тем не менее, несмотря на снижение динамики, в 2023 г.

количество российских публикаций в WoS CC больше, чем в 2014–2015 гг. (см. рис. 2).

Рассмотрим динамику российских массивов публикаций не по количеству документов, а по долям в общемировом потоке по базам данных OpenAlex и The Lens на фоне WoS CC (рис. 3). На графике видно, что доли российского массива на мировом фоне как в OpenAlex, так и в The Lens уступают WoS CC. Несмотря на лидерство открытых ресурсов по общему числу документов, первенство WoS CC по долевого распределению российского массива на общемировом фоне объясняется прежде все-

Таблица 4. Пересекающиеся предметные категории OpenAlex и The Lens, по которым доля российских публикаций на мировом фоне в The Lens выше, чем в OpenAlex, 2019–2023 гг. (нисходящая сортировка по доле российских публикаций в The Lens с минимальным порогом 1.5% мирового массива)

Предметная категория OpenAlex/The Lens	Всего публикаций в мире		Всего российских публикаций		Доля российских публикаций		Преимущество The Lens (величина отношения доли в The Lens и в OpenAlex)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Nuclear and High Energy Physics	161860	66797	9002	4554	5.6	6.8	1.2
Statistical and Nonlinear Physics	132140	38268	4661	1830	3.5	4.8	1.4
Geochemistry and Petrology	37362	71180	1211	3025	3.2	4.3	1.3
Astronomy and Astrophysics	318820	91769	12389	3833	3.9	4.2	1.1
General Decision Sciences	8855	29983	78	1189	0.9	4	4.4
Mathematical Physics	89690	42860	3987	1640	4.5	3.8	0.8
Space and Planetary Science	11029	107674	172	3717	1.6	3.5	2.2
Computational Mathematics	6065	78731	124	2301	2	2.9	1.5
Metals and Alloys	9052	182988	173	5067	1.9	2.8	1.5
Energy Engineering and Power Technology	24886	248172	291	5361	1.2	2.2	1.8
Instrumentation	18893	164037	307	3536	1.6	2.2	1.4
Modeling and Simulation	84844	144868	1336	2924	1.6	2	1.3
Statistics and Probability	102850	99356	944	1976	0.9	2	2.2
Insect Science	108690	68411	1953	1293	1.8	1.9	1.1
History	295320	556173	1844	9504	0.6	1.7	2.8
Computer Science Applications	97970	612239	548	8984	0.6	1.5	2.5
Theoretical Computer Science	12236	181871	103	2657	0.8	1.5	1.9

го выстроенной методологией отбора источников публикаций, а также контролем корректности метаданных. Начиная с 2021 г. динамика долей российских публикационных массивов замедлилась по всем трём ресурсам. Интересно, что в период 2021–2023 гг. доли российских публикаций оказались схожими в OpenAlex и WoS CC. Наименьший показатель – в The Lens, причём в этой базе порог в 1% общемирового массива был преодолен только в 2021–2022 гг. Значительная недопредставленность российских публикаций на общемировом фоне в The Lens по сравнению с OpenAlex и WoS CC

ставит под сомнение перспективы изучения российской научной публикационной активности на основе данного ресурса.

Как уже отмечалось, каждый ресурс располагает своим собственным классификатором, а методология, согласно которой публикации отнесены к тем или иным научным категориям, не всегда прозрачна и может меняться с течением времени. Например, наблюдается тенденция перехода от наиболее распространённой классификации на уровне журналов (WoS CC) к классификациям на уровне отдельных публикаций (OpenAlex) [28]. Категории в разных

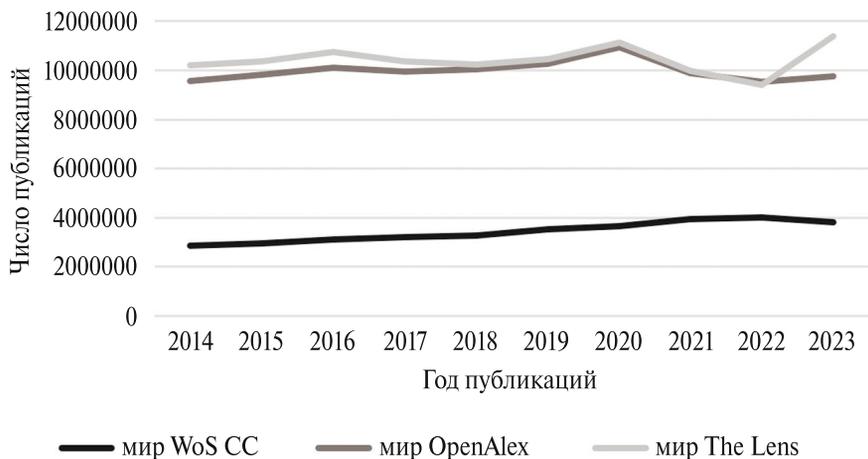


Рис. 1. Мировой поток публикаций в OpenAlex и The Lens за 2014–2023 гг. в сравнении с WoS CC

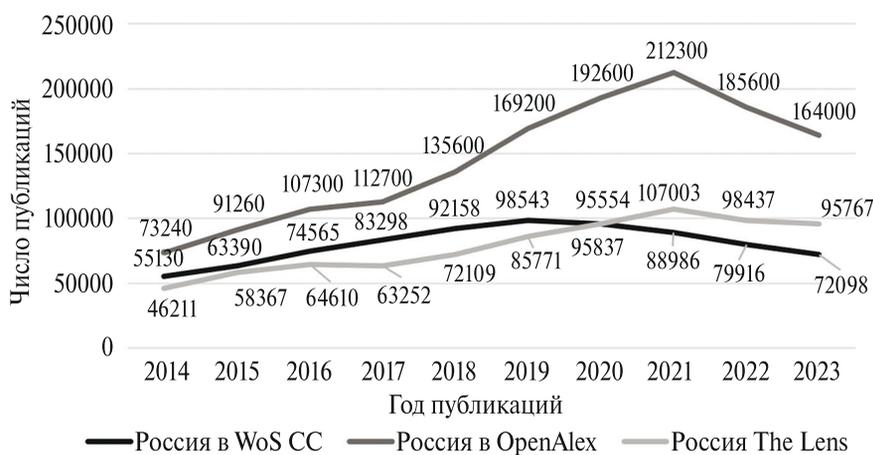


Рис. 2. Российский поток публикаций в OpenAlex и The Lens за 2014–2023 гг. в сравнении с WoS CC

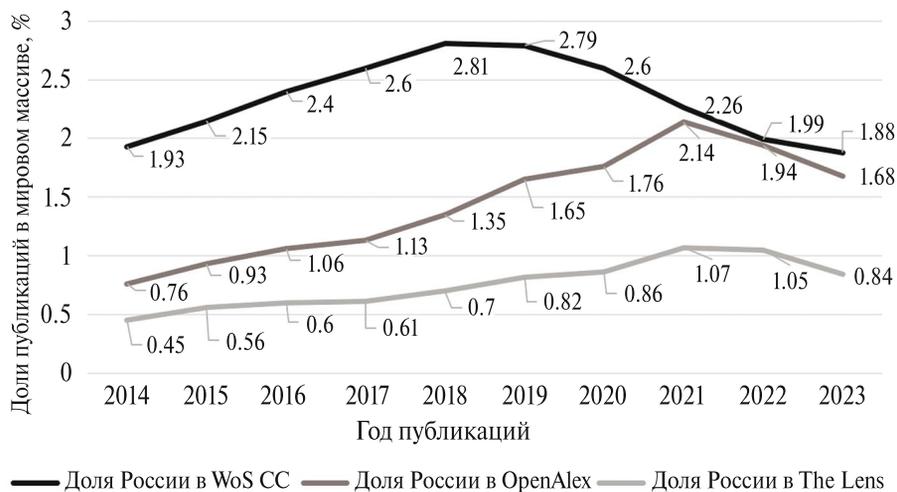


Рис. 3. Динамика доли российского сегмента публикаций в мировом массиве по OpenAlex и The Lens в сравнении с WoS CC за 2014–2023 гг.

системах могут быть как более общими, так и более частными, а одно и то же название дочерних предметных категорий может фигурировать в разных родительских категориях. Например, в OpenAlex название дочерней категории “Biochemistry” фигурирует в двух родительских – “Biochemistry, Genetics and Molecular Biology” и “Medicine”. Несмотря на то, что названия категорий в OpenAlex и The Lens в большинстве своём идентичны в пересечениях, статистика по числу публикаций в пересекающихся предметных категориях меняется от ресурса к ресурсу: по одним категориям лидирует OpenAlex, по другим – The Lens (табл. 3, 4). Если в этих ресурсах категории в большинстве своём близки, то с WoS CC таких совпадений гораздо меньше. Поэтому для сравнения ресурсов на фоне WoS CC были отобраны только наиболее близкие пересекающиеся предметные категории.

Сравним российскую публикационную активность по различным предметным категориям в OpenAlex и The Lens между собой и на фоне WoS CC.

OpenAlex на фоне WoS CC. При сравнении перечней предметных категорий в WoS CC и OpenAlex было выявлено 96 общих – 38% от 252 дочерних категорий в OpenAlex и 37% от 257 предметных категорий WoS CC. Обнаружено, что в 60 пересекающихся категориях средняя доля российских публикаций за 2014–2023 гг. в мировом массиве в WoS CC была больше, чем в OpenAlex.

Анализ динамики долей российских публикаций по пересекающимся предметным категориям в OpenAlex на фоне WoS CC показал, что в течение исследуемого периода наблюдалось сокращение разрыва между долями российских публикаций в мировом массиве в OpenAlex относительно WoS CC. Если в 2014–2018 гг. WoS CC был несомненным лидером по доле отражённого российского массива на мировом фоне, то в 2019–2023 гг. показатели этих двух ресурсов сблизались. Однако остаётся значительное число научных категорий, по которым доля российского массива на мировом фоне в WoS CC ощутимо больше, чем в OpenAlex. В таблице 1 приведены предметные категории, в которых особенно заметно превалирование WoS CC над OpenAlex. В качестве порогового значения выбрана доля публикаций в 1.5% в WoS CC, чтобы отсеять категории, в которых не наблюдалось заметной публикационной активности.

Как видим, наиболее существенные расхождения между долями российских публикаций 2019–2023 гг. в WoS CC и OpenAlex отмечены в категориях “Information Science & Library Science” и “Forestry” – в 17 и 14.5 раз соответственно. База WoS CC имела преимущество над OpenAlex в отражении российских публикаций по естественным, техническим и некоторым гуманитарным наукам. Категории, по которым преимущество на стороне OpenAlex, представлены в таблице 2.

Итак, по предметной категории “Cultural Studies” доля российских публикаций в OpenAlex опережает WoS CC более чем в 25 раз, а по “Developmental and Educational Psychology” – в 18 раз. Большинство категорий, по которым доля российских публикаций в 2019–2023 гг. была больше в OpenAlex, чем в WoS CC, – это направления медицины.

The Lens на фоне WoS CC. В The Lens и WoS CC выявлены 63 пересекающиеся предметные категории, что соответствует 39% от 161 предметной категории в The Lens и 24% от 257 категорий в WoS CC. По 61 такой категории доля российских публикаций в WoS CC оказалась больше, чем в The Lens.

Как было показано на рисунке 3, доля российских публикаций в The Lens – наименьшая по сравнению с WoS CC и OpenAlex. Тем не менее и в этом случае в 2019–2023 гг. наблюдалось некоторое сокращение долевого разрыва между российским массивом в WoS CC и The Lens практически по всем пересекающимся предметным категориям, хотя выравнивание показателей в ближайшей перспективе вряд ли достижимо.

OpenAlex и The Lens: сравнение. OpenAlex и The Lens – открытые политематические ресурсы, в которых выявлены 133 пересекающиеся одноимённые предметные категории.

Низкие долевыми показатели российских массивов в The Lens относительно WoS CC и OpenAlex (см. рис. 3) выглядят достаточно странно, имея в виду, что количество мировых публикаций в The Lens и OpenAlex практически идентично. Слабая представленность российского документопотока в The Lens может быть вызвана такими причинами, как: ошибки в системе, вызывающие утерю метаданных в процессе индексации документов, содержащих сведения о стране; специфика формируемого контента; тенденциозность в отношении российского сегмента документов.

Исследование показало, что в OpenAlex в 106 пересекающихся предметных категориях доля российских публикаций на мировом фоне оказалась больше, чем в The Lens (табл. 3). По 43 категориям преимущество OpenAlex составляло 3 и более раз, по 17 категориям – 2 раза. Особенно выделяются категории “Agronomy and Crop Science” (превышение больше чем в 22 раза), “Food Science и General Materials Science” (в 12 раз). По 27 категориям The Lens опережает OpenAlex (табл. 4).

Таким образом, лишь по небольшому кругу предметных категорий в технических науках и науках о жизни доля российских публикаций в мировом массиве 2019–2023 гг. оказалась выше в The Lens, чем в OpenAlex. Лидерство The Lens (см. табл. 4) выглядит не столь убедительно, как в случае с OpenAlex (см. табл. 3): максимальная разница по долям в четырнадцать категорий в The Lens составляет одно-двукратное превышение над OpenAlex, и лишь в двух случаях – трёх- и четырёхкратное.

* * *

В ходе исследования обнаружено, что OpenAlex обладает преимуществами относительно другого открытого ресурса – The Lens. Несмотря на высокие абсолютные показатели динамики мирового потока в обеих этих базах данных, российские публикации в The Lens сильно недопредставлены. Это наглядно показал анализ долевого распределения российских публикаций на общемировом фоне и по большинству пересекающихся предметных категорий.

На текущий момент лидером среди рассмотренных трёх ресурсов по долевному соотношению российского и мирового массивов публикаций остаётся WoS CC. Тем не менее начиная с 2019 г. наблюдается активное сокращение разрыва в долях между российским и мировым массивами публикаций в WoS CC, OpenAlex и The Lens. С 2021 г. данные о динамике российского массива на мировом фоне по OpenAlex коррелируют с показателями WoS CC, что позволяет надеяться на высокую релевантность результатов поиска посредством OpenAlex.

ЛИТЕРАТУРА

1. *Pranckutė R.* Web of science (WoS) and Scopus: The titans of bibliographic information in today's academic world // *Publications*. 2021, vol. 9, no. 1, art. 12. DOI: 10.3390/publications9010012
2. *Waltman L., van Eck N., Visser M.* Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic // *Quantitative Science Studies*. Advance Publication. 2021. https://doi.org/10.1162/qss_a_00112
3. *Vera-Baceta M.-A., Thelwall M., Kousha K.* Web of Science and Scopus language coverage // *Scientometrics*. 2019, vol. 121, pp. 1803–1813. <https://doi.org/10.1007/s11192-019-03264-z>
4. *Mongeon P., Paul-Hus A.* The journal coverage of Web of Science and Scopus: a comparative analysis // *Scientometrics*. 2016, vol. 106, pp. 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
5. *Мохначева Ю.В., Цветкова В.А.* Возможные пути получения научной информации в новых условиях // *Управление наукой: теория и практика*. 2023. Т. 5. № 3. С. 117–158. DOI 10.19181/sntp.2023.5.3.9
Mokhnacheva Yu.V., Tsvetkova V.A. Possible Ways to Search for Scientific Information in New Realities // *Science Management: Theory and Practice*. 2023, vol. 5, no. 3, pp. 117–158. DOI: <https://doi.org/10.19181/sntp.2023.5.3.9> (In Russ.)
6. *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/sntp.2023.5.2.4
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, vol. 5, no. 2, pp. 49–76. DOI: <https://doi.org/10.19181/sntp.2023.5.2.4> (In Russ.)
7. В странах БРИКС намерены создать единую наукометрическую базу данных. <https://nauka.tass.ru/nauka/21976405>
The BRICS countries intend to create a single scientometric database. <https://nauka.tass.ru/nauka/21976405>
8. *Delgado-Quirós L., Aguillo I.F., Martín-Martín A. et al.* Why are these publications missing? Uncovering the reasons behind the exclusion of documents in free-access scholarly databases // *Journal of the Association for Information Science and Technology*. 2024, vol. 75, pp. 43–58. <https://doi.org/10.1002/asi.24839>
9. *Мохначева Ю.В.* Типы документов, индексируемых в базах данных WoS и Scopus: сходства, различия и их значение при анализе публикационной активности // *Научно-техническая информация. Серия 1: Организация и методика информационной работы*. 2023. № 1. С. 38–43. DOI: 10.36535/0548-0019-2023-01-4
Mokhnacheva Yu.V. Document Types Indexed in WoS and Scopus: Similarities, Differences, and Their Significance in the Analysis of Publication Activity // *Sci. Tech. Inf. Proc.* 2023, vol. 50, pp. 40–46. <https://doi.org/10.3103/S0147688223010033> (In Russ.)
10. *Donner P.* Document type assignment accuracy in the journal citation index data of Web of Science // *Scientometrics*. 2017, vol. 113, pp. 219–236. DOI: 10.1007/s11192-017-2483-y
11. *Blümel C., Schniedermann A.* Studying review articles in scientometrics and beyond: a research agenda // *Scientometrics*. 2020, vol. 124, pp. 711–728. DOI: 10.1007/s11192-020-03431-7
12. *Gerasimov I., K.B., Mehrabian A., Acker J., Mcguire M.P.* Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite // *Scientometrics*. 2024, vol. 129, pp. 3681–3704. <https://doi.org/10.1007/s11192-024-05073-5>
13. *Мальцева Д.В., Павлова И.А., Капустина Л.В.* Сравнительный анализ возможностей WoS и eLibrary для анализа библиографических сетей // *Социология: 4М*. 2023. № 56. С. 7–68. <https://doi.org/10.19181/4m.2023.32.1.1>
14. *Maltseva D., Pavlova I., Kapustina L.* Comparative analysis of the capabilities of WoS and eLibrary for analyzing bibliographic networks // *Sociology: methodology, methods, mathematical modeling (Sociology: 4M)*. 2023, no. 56, pp. 7–68. <https://doi.org/10.19181/4m.2023.32.1.1> (In Russ.)

15. *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. С. 20–32. DOI: 10.36535/0548-0019-2023-09-4
16. *Гуреев В.Н., Ильичёва И.Ю., Мазов Н.А.* Профили авторов и организаций в информационных системах Dimensions и Lens: исследование возможностей // Научные и технические библиотеки. 2023. № 10. С. 138–170. <https://doi.org/10.33186/1027-3689-2023-10-138-170>
Gureyev V.N., Illicheva I.Yu., Mazov N.A. Author and organization profiles in Dimensions and Lens information systems: The study of functionality // Scientific and Technical Libraries. 2023, no. 10, pp. 138–170. <https://doi.org/10.33186/1027-3689-2023-10-138-170> (In Russ.)
17. *Tay A., Martín-Martín A., Hug S.E.* Goodbye, Microsoft Academic – hello, open research infrastructure? // Impact of Social Sciences Blog. 27 May 2021. <http://eprints.lse.ac.uk/id/eprint/111325>
18. Next Steps for Microsoft Academic – Expanding into New Horizons // Microsoft Academic Blog. May 4, 2021. <https://www.microsoft.com/en-us/research/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach>
19. *Chawla D.S.* Microsoft Academic Graph is being discontinued. What's next? // Nature. 2021, 15 June 2021.
20. *Aria M., Le T., Cuccurullo C. et al.* openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex // The R Journal. 2024, vol. 15, pp. 167–180. <https://doi.org/10.32614/rj-2023-089>
21. *Priem J., Piwowar Y., Orr R.* OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts // STI Conference 2022, Granada. arXiv:2205.01833. <https://doi.org/10.48550/arXiv.2205.01833>
22. *Chawla D.S.* Massive open index of scholarly papers launches. OpenAlex catalogues hundreds of millions of scientific documents and charts connections between them // Nature. 2022. 24 January. <https://www.nature.com/articles/d41586-022-00138-y>
23. OpenAlex. Техническая документация. <https://docs.openalex.org/api-entities/works>
24. *Brooks J.* Leiden rankings to add open-source version in 2024: Research Professional News. <https://www.researchprofessionalnews.com/rr-news-europe-universities-2023-9-leiden-rankings-to-add-open-source-version-in-2024/>
25. CWTS Leiden Ranking Open Edition. <https://open.leidenranking.com/>
26. *Penfold R.* Using the Lens database for staff publications // Journal of the Medical Library Association. 2020, vol. 108. <https://doi.org/10.5195/jmla.2020.918>
27. *Jefferson O.A., Koellhofer D., Warren B., Jefferson R.* The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts. 2019, November 25. <https://doi.org/10.31229/osf.io/t56yh>
28. The Lens: статья из Википедии. https://en.wikipedia.org/wiki/The_Lens
29. OpenAlex: End-to-End Process for Topic Classification. <https://docs.google.com/document/d/1bDopkhuGieQ4F8gGNj7sEc8WSE8mvLZS/edit#heading=h.5w2tb5fcg77r>

RUSSIA IN THE GLOBAL ARRAY OF SCIENTIFIC PUBLICATIONS**Yu.V. Mokhnacheva^{a,*}***^aLibrary for Natural Sciences RAS, Moscow, Russia***E-mail: j-v-m@yandex.ru*

Selecting the most relevant platform for monitoring the Russian publication array against the global background, including the analysis of publication flows in various scientific categories, is a pressing task. The objects of the study were the Russian and global publication arrays in the OpenAlex and The Lens databases in comparison with the Web of Science Core Collection (WoS CC) for the period 2014–2023, with a special emphasis on the period 2019–2023. It was found that OpenAlex has a number of advantages over The Lens, since Russian publications are significantly underrepresented in The Lens. A comparative analysis of the dynamics of the shares of Russian publications by subject categories in OpenAlex and The Lens against the background of WoS CC showed that during 2019–2023, the gap between the shares of Russian publications in the global array in these resources relative to WoS CC for most intersecting subject categories narrowed. The data on the share distribution of Russian publications in 2021–2023 in OpenAlex correlate with the indicators in WoS CC, which gives hope for high relevance of search results using OpenAlex.

Keywords: science in Russia, Russia's contribution to world science, open information and bibliographic databases, arrays of publications, OpenAlex, The Lens, scientometric research, subject categories.