HERALD OF THE RUSSIAN ACADEMY OF SCIENCES, 2025, Vol. 95, No. 1, pp. 42–55

ORGANIZATION OF RESEARCH

RUSSIA IN THE GLOBAL ARRAY OF SCIENTIFIC PUBLICATIONS

© 2025 Yu. V. Mokhnacheva*

Library for Natural Sciences, Russian Academy of Sciences, Moscow, Russia
**e-mail: j-v-m@yandex.ru*

Received October 03, 2024
Revised October 29, 2024
Accepted December 02, 2024

Abstract. 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*

DOI: 10.31857/S08695873250105e2

Since the introduction of unprecedented anti-Russian sanctions by the United States, the European Union and other unfriendly countries, Russian users have been indefinitely denied access to the most well-known and reputable scientific, technical and scientometric information search resources: Web of Science (Clarivate, the USA) and Scopus (Elsevier, the Netherlands). Despite some disadvantages of these databases [1–4], they are used to generate reports on publication activity at all levels: micro (the level of individual personalities and publications), meso (the level of organizations), and macro (the level of states). With the withdrawal of these databases from the Russian information space, the task of finding alternative ways to obtain information about publication activity, especially at the macro level, has become particularly acute.

MOKHNACHEVA Yulia V., Candidate of Pedagogical Sciences, leading researcher at the Library for Natural Sciences, RAS.

To implement a successful scientific policy, relevant data on the state and dynamics of Russian publication arrays against a global background is needed, including information on individual scientific areas, priority ones among them. The authority of the Web of Science (WoS) and Scopus is formed mainly due to the wide functionality and high-quality content: the mere fact that journals are included in these databases implies high quality sources, since it is based on strict selection rules [1]. Nevertheless, the time has come to find alternatives to these systems.

Although users have at their disposal a fairly wide range of open resources (OpenAlex, The Lens, Dimensions, Scilit, Semantic Scholar, RSCI, CoLab, etc.) that allow them to obtain diverse information about publications, including bibliometric and altmetric data [5, 6], not all systems meet the necessary requirements and have full functionality. The possibilities of such resources are often limited by information about publications at the level of authors and their publications, less often at the level of organizations and very rarely at the level of states.

Significant progress has been made in the creation and development of bibliographic and scientometric systems in Russia. Databases are successfully operating and actively developing such as the RSCI on the eLibrary.ru platform (Scientific Electronic Library LLC)[1], CoLab[2], CyberLeninka[3], the portal "White List" of journals (RCSI)[4], the ISTINA system (Lomonosov Moscow State University)[5], etc. On September 27, 2024, V.N. Falkov, Minister of Science and Higher Education of the Russian Federation, announced the creation of a unified scientometric database of the BRICS countries [7]. Despite the wide range of open Russian and foreign bibliographic resources, only two of them – OpenAlex and The Lens – allow us to determine Russia's status in the global scientific space.

Open or partially open access resources have obvious benefits and advantages: they are free of charge, versatile, polythematic, and wide-ranging in global scientific content. Bibliographic information resources based on search engines and web scanning, including open ones, provide more comprehensive document coverage than WoS and Scopus [8]. Nevertheless, a number of questions are raised about such systems: how high is the authority of indexed sources; how correct is the metadata; how transparent is the policy and methodology for selecting and indexing sources, etc.? Inconsistencies in the metadata of publications in bibliographic resources negatively affect the relevance and accuracy of the results obtained [1]. To be fair, it should be noted that the "reference" WoS and Scopus are also not without drawbacks [9–11].

Comparing different databases with each other is an independent area of research in the information and library field. Resources are evaluated according to various criteria, from the coverage of indexed publications and the depth of database retrospect to differences in the methodology of document typification and subject classification, as well as the possibilities of application in information and bibliographic activities [1, 2, 6, 9–15].

In order to compare the coverage of various scientific databases, based on the principle of random sampling from Crossref, the authors of the publication [8] suggested that using a third-party resource to compare the coverage of the other two would reduce possible bias and make it possible to find out how the selection criteria and technical requirements affect the coverage of scientific literature. The researchers concluded that from the perspective of Crossref, there are significant differences in coverage between scientific databases, and they can usually be explained by differences in methodology when creating databases, including document typification. In [2], five multidisciplinary bibliographic resources for the period 2008–2017 were compared: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Each of the systems was compared with Scopus. The authors focus on the problems associated with the classification of document types and note the strong dependence of Dimensions on data from Crossref.

Errors in author affiliations are a common significant problem of proprietary resources and open access systems [2]. Such inaccuracies directly affect the statistics of publication activity at the meso and macro levels; they probably arise at the stage of indexing publications by information systems. All bibliographic database providers, without exception, are concerned about this problem, but it has not yet been resolved.

In 2015–2021, the Microsoft Academic open resource enjoyed well-deserved popularity. According to the conclusions presented in the article [2], this database covered and indexed a wider range of documents than other data sources. The predominant share was accounted for by scientific papers, which was proved by the results of manual sampling. Microsoft Academic's strongest point was the technology of scanning web pages, followed by the use of artificial intelligence to update the database with comprehensive metadata, which allowed for faster indexing of new documents [16]. Another advantage of Microsoft Academic was the large number of reflected documents that are not in English. In the process of searching for new ways to empower researchers, the project was implemented in two technological solutions: first, all web pages indexed by Bing were read, the most up-to-date scientific information was selected, and a database called the Microsoft Academic Graph (MAG)[6] was generated; then semantic justification and inference were carried out to serve this knowledge through a search on the Microsoft Academic website [17]. The developers' decision to discontinue support for Microsoft Academic in May and MAG in December 2021 caused serious concern among users [18].

In 2022, as an alternative to Microsoft Academic and MAG, the American company OurResearch[7] launched the OpenAlex system, named after the ancient library of Alexandria. OpenAlex[8] is a free and fully open-source catalog of scientific metadata that allows for comprehensive bibliographic analysis. Like Google

[1] RSCI. https://elibrary.ru/defaultx.asp?
[2] CoLab. https://colab.ws/
[3] CyberLeninka scientific electronic library. https://cyberleninka.ru/
[4] "White List" of scientific journals, RCSI. https://journalrank.rcsi.science/ru/
[5] ISTINA. https://istina.msu.ru/
[6] Microsoft Academic Graph (MAG). https://www.microsoft.com/en-us/research/project/academic/
[7] It is known in the scientific and publishing community for the Unpaywall database and plugin.
[8] OpenAlex. https://openalex.org/works

Scholar, OpenAlex does not use expert assessments when selecting sources, preferring the breadth of content coverage [19–21].

OpenAlex collects and standardizes data from many sources, primarily from MAG and Crossref[9], as well as data from ORCID[10], ROR[11], DOAJ[12], Unpaywall[13], Pubmed[14], Pubmed Central[15], The ISSN International Centre[16], and various repositories. OpenAlex indexes more than 240 million papers, and about 50,000 documents are added daily [21, 22]. In 2024, the University of Leiden (the Netherlands) used OpenAlex to compile a rating of research institutions [23, 24]. OpenAlex makes it possible to download the necessary data both via the API and by downloading data from the search results page in CSV format. An important advantage of the system is the ability to obtain information about publication arrays at all levels: micro, meso, and macro.

The Lens[17] database was created in 1998. It is a free platform for searching scientific and patent literature, a joint project of the non-profit organization Cambia and the Queensland University of Technology (Australia) [25]. It implements the concept of MetaRecord (MeR), which allows managing problems related to the variability of records, sources, and the contextual relevance of metadata to the original record [26]. The Lens uses information from third-party systems (such as PubMed and Crossref), combining them into one database deduplicated with a unified search syntax. Unlike competing databases, the resource allows you to export data in the JSON format with more thorough details compared to the RIS and CSV formats [27].

Currently, OpenAlex and The Lens are, in fact, the only open resources through which it is possible to find and download data not only about individual publications, authors and scientific organizations, but also about publication arrays at the macro level.

Research methodology. The object of the study was the Russian array of publications in the databases OpenAlex and The Lens in comparison with WoS CC. The purpose of the study is to select the optimal system for monitoring the Russian array of publications, as well as to determine the share of Russian publications in the global scientific and information field using the OpenAlex and The Lens databases in comparison with the WoS CC data for 2014–2023 (emphasis on 2019–2023).

In accordance with the statement that referring to a third-party resource to compare the coverage of the other two reduces possible biases [2], data on the dynamics of the Russian array of publications against the global background in WoS CC were taken as reference. The open WoS CC module was used, which allows searching[18] without the possibility of uploading information to a file. Data was collected for each of the three resources, WoS CC, OpenAlex, and The Lens for the world in general and for Russia in particular for each year of the period 2014–2023, including overlapping subject categories.

To understand the dynamic characteristics of the Russian document flow, given the actively narrowing range of resources for both authors of publications, as well as for specialists in the field of scientometry and scientific administrators, the time interval 2019–2023 is especially important. Therefore, the study was conducted with an emphasis on this period.

Certain difficulties were associated with the fact that databases use their own classifiers, which means that not all the names of subject categories in one resource are completely identical to the categories in another. Therefore, the results of comparing databases are presented here only in overlapping scientific categories (Tables 1–4).

When collecting information, all types of documents were taken into account, since it was noted that from resource to resource the same works can relate to different types of documents [9–11]. In WoS CC, the search was conducted simultaneously for all indexes as of July–August 2024: 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.

Features of information representation in OpenAlex and The Lens. To understand the breadth of content coverage in OpenAlex and The Lens against the background of WoS CC, Fig. 1 shows the dynamics of the number of publications in the world for the period 2014–2023 (as of July–August 2024). As you can see, in terms of the number of publications, the open resources of OpenAlex and The Lens are many times higher than the statistics for WoS CC. At the same time, with the exception of 2023, the dynamics of publications in OpenAlex and The Lens are almost identical, which suggests a possible high degree of similarity in the content of these resources. However, this assumption is refuted both by the results of the study of the dynamics of the Russian

[9] Crossref. https://www.crossref.org/
[10] Open Researcher and Contributor ID. 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:// URL: www.lens.org/

[18] The country search option is available in the open WoS CC module by clicking on the active links provided in the WoS CC publication card (subject to registration and authorization in the system). Further, by activating the search field, it is possible to access the extended search menu for all fields.

Table 1. Overlapping subject categories of WoS CC and OpenAlex, for which the share of Russian publications against the global background in WoS CC is higher than in OpenAlex, 2019–2023 (top-down sorting by the share of Russian publications in WoS CC with a minimum threshold of 1.5% of the global array)

Subject category		Total number of publications worldwide in the period		Total number of Russian publications in the period		Share of Russian publications, %		Advantage of the share of the Russian array in WoS CC over OpenAlex (value of the share ratio in WoS CC and in OpenAlex)
WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	
Physics, Nuclear	Nuclear and High Energy Physics	39,571	161,860	4,784	9,002	12.1	5.6	2.2
Paleontology	Paleontology	19,367	53,286	1,767	1,807	9.1	3.4	2.7
Chemistry, Inorganic & Nuclear	Inorganic Chemistry	75,888	105,910	5,622	3,848	7.4	3.6	2.1
Physics, Mathematical	Mathematical Physics	65,239	89,690	4,773	3,987	7.3	4.5	1.6
Astronomy & Astrophysics	Astronomy and Astrophysics	135,094	318,820	9,596	12,389	7.1	3.9	1.8
Spectroscopy	Spectroscopy	40,176	89,490	2,571	3,478	6.4	3.9	1.6
Chemistry, Organic	Organic Chemistry	100,573	325,480	5,469	10,948	5.4	3.4	1.6
Physics, Condensed Matter	Condensed Matter Physics	236,135	79,510	12,816	3,236	5.4	4.1	1.3
Oceanography	Oceanography	59,583	161,690	3,197	6,039	5.4	3.7	1.5
Mathematics, Applied	Applied Mathematics	227,173	146,230	10,374	4,905	4.6	3.4	1.4
History	History	272,941	295,320	11,697	1,844	4.3	0.6	7.2
Archaeology	Archeology	43,272	249,140	1,782	3,029	4.1	1.2	3.4
Soil Science	Soil Science	38,871	140,690	1,597	4,061	4.1	2.9	1.4
Language & Linguistics	Language and Linguistics	97,338	283,330	3,945	5,389	4.1	1.9	2.2
Zoology	Animal Science and Zoology	106,204	85,850	4,111	1,148	3.9	1.3	3
Ecology	Ecology	148,089	360,180	5,331	7,767	3.6	2.2	1.6

Table 1. (Continued)

Subject category		Total number of publications worldwide in the period		Total number of Russian publications in the period		Share of Russian publications, %		Advantage of the share of the Russian array in WoS CC over OpenAlex (value of the share ratio in WoS CC and in OpenAlex)
WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	
Economics	Economics and Econometrics	265,931	808,200	8,589	11,642	3.2	1.4	2.3
Instruments & Instrumentation	Instrumentation	182,888	18,893	5,573	307	3.1	1.6	1.9
Philosophy	Philosophy	142,785	272,840	4,235	1,998	3	0.7	4.3
Forestry	Forestry	47,444	62,620	1,359	142	2.9	0.2	14.5
Sociology	Sociology and Political Science	103,373	2,020,900	2,790	25,932	2.7	1.3	2.1
Education & Educational Research	Education	287,957	1,318,600	7,510	11,164	2.6	0.9	2.9
History & Philosophy of Science	History and Philosophy of Science	44,167	133,520	1,053	1,016	2.4	0.8	3
Physiology	Physiology	82,140	348,440	1,782	5,233	2.2	1.5	1.5
Information Science & Library Science	Library and Information Sciences	72,565	34,203	1,204	46	1.7	0.1	17
Anthropology	Anthropology	60,022	233,930	974	1,969	1.6	0.8	2
Developmental Biology	Developmental Biology	32,300	10,103	502	120	1.6	1.2	1.3
Transplantation	Transplantation	73,582	14,214	1,103	114	1.5	0.8	1.9

array of publications (Fig. 2) and the analysis of publication arrays by overlapping subject categories (Tables 3, 4).

Fig. 2 shows that OpenAlex is the leader in indexing Russian publications for each year of the 2014–2023 period, and this is especially noticeable after 2017. The decrease in dynamics in the last two years, in contrast to the active acceleration in 2017–2021, typical for all three systems, is apparently due to technical delays in indexing documents. Until 2019, WoS CC was ahead of The Lens in terms of the number of Russian documents, but since 2020, the growth of the Russian array of publications in WoS CC has noticeably slowed down. Nevertheless, despite the decrease in dynamics, in 2023 the number of Russian publications in WoS CC was higher than in 2014–2015 (see Fig. 2).

Table 2. Overlapping subject categories of WoS CC and OpenAlex, for which the share of Russian publications against the global background in OpenAlex is higher than in WoS CC, 2019–2023 (top-down sorting by the share of Russian publications in OpenAlex with a minimum threshold of 1.5% of the global array)

Subject category		Total number of publications worldwide in the period		Total number of Russian publications in the period		Share of Russian publications		Advantage of the share of the Russian array in OpenAlex (value of the share ratio in OpenAlex and in WoS CC)
OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	OpenAlex	WoS CC	
Development	Development Studies	76,750	33,354	11,739	1,254	15.3	3.8	4
Anatomy	Anatomy & Morphology	21,441	19,248	1,832	307	8.5	1.6	5.3
Agronomy and Crop Science	Agronomy	109,730	102,145	7,289	1,622	6.6	1.6	4.1
Cultural Studies	Cultural Studies	248,900	52,865	12,721	110	5.1	0.2	25.5
Mechanical Engineering	Engineering, Mechanical	625,200	237,711	29,622	7,167	4.7	3	1.6
Acoustics and Ultrasonics	Acoustics	5,340	52,864	224	1,191	4.2	2.3	1.8
Food Science	Food Science & Technology	302,160	243,197	11,120	3,170	3.7	1.3	2.8
Developmental and Educational Psychology	Psychology, Developmental	179,690	59,895	6,498	119	3.6	0.2	18
Demography	Demography	306,280	20,364	10,718	189	3.5	0.9	3.9
Analytical Chemistry	Chemistry, Analytical	68,950	202,464	2,162	4,248	3.1	2.1	1.5
Toxicology	Toxicology	15,977	98,343	352	755	2.2	0.8	2.8
Transportation	Transportation	81,910	94,936	1,793	911	2.2	1	2.2
Pediatrics, Perinatology and Child Health	Pediatrics	317,690	210,115	6,276	973	2	0.5	4
Ophthalmology	Ophthalmology	113,080	121,952	1,995	478	1.8	0.4	4.5
Radiology, Nuclear Medicine and Imaging	Radiology, Nuclear Medicine & Medical Imaging	374,070	246,871	6,430	1,551	1.7	0.6	2.8
Urology	Urology & Nephrology	48,136	147,802	829	1,476	1.7	1	1.7
Immunology	Immunology	238,860	255,004	4,044	2,615	1.7	1	1.7
Infectious Diseases	Infectious Diseases	337,000	146,706	4,942	1,363	1.5	0.9	1.7

Table 3. Overlapping subject categories of OpenAlex and The Lens, for which the share of Russian publications in OpenAlex is higher than in The Lens, compared to the global background, 2019–2023 (top-down sorting by the share of Russian publications in OpenAlex with a minimum threshold of 1.5% of the global array)

Subject category OpenAlex/The Lens	Total number of publications worldwide		Total number of Russian publications		Share of Russian publications		Advantage of the share of the Russian array in OpenAlex over the Lens (value of the share ratio in OpenAlex and in The Lens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Fuel Technology	6,338	135,808	1,510	3,502	23.8	2.6	9.2
Nuclear Energy and Engineering	3,843	58,093	876	1,651	22.8	2.8	8.1
General Materials Science	32,629	753,923	6,509	12,957	20	1.7	11.8
Geology	71,610	108,917	5,670	2,661	7.9	2.4	3.3
Mechanics of Materials	267,620	367,260	19,429	8,118	7.3	2.2	3.3
Agronomy and Crop Science	109,730	193,636	7,289	667	6.6	0.3	22
Earth-Surface Processes	48,576	88,055	3,202	2,083	6.6	2.4	2.8
Ceramics and Composites	24,267	123,200	1,517	2,839	6.3	2.3	2.7
General Energy	16,274	96,058	991	1,082	6.1	1.1	5.5
General Economics, Econometrics and Finance	168,900	75,674	9,897	1,052	5.9	1.4	4.2
Atomic and Molecular Physics, and Optics	297,970	292,242	16,919	9,567	5.7	3.3	1.7
Cultural Studies	248,900	237,453	12,721	2,712	5.1	1.1	4.6
Industrial and Manufacturing Engineering	205,360	521,695	10,336	7,384	5	1.4	3.6
Mechanical Engineering	625,200	540,833	29,622	9,574	4.7	1.8	2.6
Catalysis	41,055	191,677	1,691	2,883	4.1	1.5	2.7
Atmospheric Science	204,490	83,959	8,361	1,807	4.1	2.2	1.9
Spectroscopy	89,490	156,072	3,478	3,158	3.9	2	2
Physiology (Medicine)	18,956	129,407	730	946	3.9	0.7	5.6
Food Science	302,160	259,669	11,120	875	3.7	0.3	12.3
Organic Chemistry	325,480	356,484	10,948	6,528	3.4	1.8	1.9
Control and Systems Engineering	363,210	219,988	11,488	2,698	3.2	1.2	2.7
Analytical Chemistry	68,950	20,8691	2,162	2,212	3.1	1.1	2.8

Table 3. (Continued)

Subject category OpenAlex/The Lens	Total number of publications worldwide		Total number of Russian publications		Share of Russian publications		Advantage of the share of the Russian array in OpenAlex over the Lens (value of the share ratio in OpenAlex and in The Lens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Biophysics	48,194	122,030	1,467	1,867	3	1.5	2
Information Systems	809,600	148,238	24,493	1,199	3	0.8	3.8
Soil Science	140,690	77,988	4,061	1,056	2.9	1.4	2.1
Neurology (Medicine)	97,190	203,879	2,724	801	2.8	0.4	7
Ecology, Evolution, Behavior and Systematics	236,010	312,256	6,554	4,615	2.8	1.5	1.9
Biomedical Engineering	694,200	163,838	17,959	1,631	2.6	1	2.6
Polymers and Plastics	116,770	321,306	3,016	4,177	2.6	1.3	2
Strategy and Management	400,480	214,984	10,092	828	2.5	0.4	6.3
Management, Monitoring, Policy and Law	305,050	225,423	7,648	1,636	2.5	0.7	3.6
Electrochemistry	16,220	76,489	403	907	2.5	1.2	2.1
Pharmaceutical Science	45,256	240,888	1,123	1,875	2.5	0.8	3.1
Electrical and Electronic Engineering	1,165,300	771,632	28,597	8,295	2.5	1.1	2.3
General Social Sciences	72,816	82,197	1,758	816	2.4	1	2.4
Law	404,280	153,243	9,684	1,216	2.4	0.8	3
Bioengineering	15,572	138,816	362	1,133	2.3	0.8	2.9
Building and Construction	233,360	186,502	5,405	1,062	2.3	0.6	3.8
Computational Theory and Mathematics	194,420	76,846	4,509	654	2.3	0.9	2.6
Ecology	360,180	202,313	7,767	2,312	2.2	1.1	2
Cardiology and Cardiovascular Medicine	389,200	434,927	7,797	1,751	2	0.4	5
Pediatrics, Perinatology and Child Health	317,690	321,291	6,276	690	2	0.2	10
Cellular and Molecular Neuroscience	106,820	106,528	2,108	501	2	0.5	4

Table 3. (Continued)

Subject category OpenAlex/The Lens	Total number of publications worldwide		Total number of Russian publications		Share of Russian publications		Advantage of the share of the Russian array in OpenAlex over the Lens (value of the share ratio in OpenAlex and in The Lens)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Biochemistry	33,775	616,063	648	6,331	1.9	1	1.9
Physical Therapy, Sports Therapy and Rehabilitation	94,410	116,473	1,816	385	1.9	0.3	6.3
Civil and Structural Engineering	423,680	238,539	7,950	898	1.9	0.4	4.8
Genetics	353,910	309,737	6,601	2,090	1.9	0.7	2.7
Internal Medicine	29,023	116,339	536	355	1.9	0.3	6.3
Microbiology	44,600	151,286	824	1,129	1.9	0.8	2.4
Renewable Energy, Sustainability and the Environment	280,460	319,956	5,191	1,902	1.9	0.6	3.2
Ophthalmology	113,080	112,170	1,995	392	1.8	0.4	4.5
Immunology and Allergy	35,945	232,002	625	1,430	1.7	0.6	2.8
Gastroenterology	46,920	183,238	798	388	1.7	0.2	8.5
Immunology	238,860	282,911	4,044	1,766	1.7	0.6	2.8
Molecular Biology	1,220,200	501,775	19,438	3,733	1.6	0.7	2.3
Biotechnology	58,670	233,457	929	1,456	1.6	0.6	2.7
Environmental Engineering	180,530	208,640	2,783	1,946	1.5	0.9	1.7
Physiology	348,440	156,046	5,233	1,305	1.5	0.8	1.9
Computer Networks and Communications	359,870	230,365	5,363	1,018	1.5	0.4	3.8
Infectious Diseases	337,000	257,944	4,942	1,099	1.5	0.4	3.8

Let us consider the dynamics of Russian publication arrays not by the number of documents, but by shares in the global flow of OpenAlex and The Lens databases against the background of WoS CC (Fig. 3). The graph shows that the shares of the Russian array against the global background in both OpenAlex and The Lens are inferior to WoS CC. Despite the leadership of open resources in terms of the total number of documents, the superiority of WoS CC in terms of the shared distribution of the Russian array against a global background is primarily due to the built-in methodology for selecting publication sources, as well as monitoring the correctness of metadata. Starting from 2021, the dynamics of the shares of Russian publication arrays has slowed down for all three resources. Interestingly, in the period 2021–2023, the

Table 4. Overlapping subject categories of OpenAlex and The Lens, for which the share of Russian publications on a global background in The Lens is higher than in OpenAlex, 2019–2023 (top-down sorting by the share of Russian publications in The Lens with a minimum threshold of 1.5% of the global array)

Subject category OpenAlex/The Lens	Total number of publications worldwide		Total number of Russian publications		Share of Russian publications		Advantage of the Lens (value of the share ratio in The Lens and in OpenAlex)
	OpenAlex	The Lens	OpenAlex	The Lens	OpenAlex	The Lens	
Nuclear and High Energy Physics	161,860	66,797	9,002	4,554	5.6	6.8	1.2
Statistical and Nonlinear Physics	132,140	38,268	4,661	1,830	3.5	4.8	1.4
Geochemistry and Petrology	37,362	71,180	1,211	3,025	3.2	4.3	1.3
Astronomy and Astrophysics	318,820	91,769	12,389	3,833	3.9	4.2	1.1
General Decision Sciences	8,855	29,983	78	1,189	0.9	4	4.4
Mathematical Physics	89,690	42,860	3,987	1,640	4.5	3.8	0.8
Space and Planetary Science	11,029	107,674	172	3,717	1.6	3.5	2.2
Computational Mathematics	6,065	78,731	124	2,301	2	2.9	1.5
Metals and Alloys	9,052	182,988	173	5,067	1.9	2.8	1.5
Energy Engineering and Power Technology	24,886	248,172	291	5,361	1.2	2.2	1.8
Instrumentation	18,893	164,037	307	3,536	1.6	2.2	1.4
Modeling and Simulation	84,844	144,868	1,336	2,924	1.6	2	1.3
Statistics and Probability	102,850	99,356	944	1,976	0.9	2	2.2
Insect Science	108,690	68,411	1,953	1,293	1.8	1.9	1.1
History	295,320	556,173	1,844	9,504	0.6	1.7	2.8
Computer Science Applications	97,970	612,239	548	8,984	0.6	1.5	2.5
Theoretical Computer Science	12,236	181,871	103	2,657	0.8	1.5	1.9

shares of Russian publications turned out to be similar in OpenAlex and WoS CC. The lowest figure is in The Lens, and in this database the threshold of 1% of the global array was overcome only in 2021–2022. The significant underrepresentation of Russian publications on a global background in The Lens compared to OpenAlex and WoS CC casts doubt on the prospects for studying Russian scientific publication activity based on this resource.

As already noted, each resource has its own classifier, and the methodology according to which publications are classified into certain scientific categories is not always transparent and may change over time. For example, there is a tendency to switch

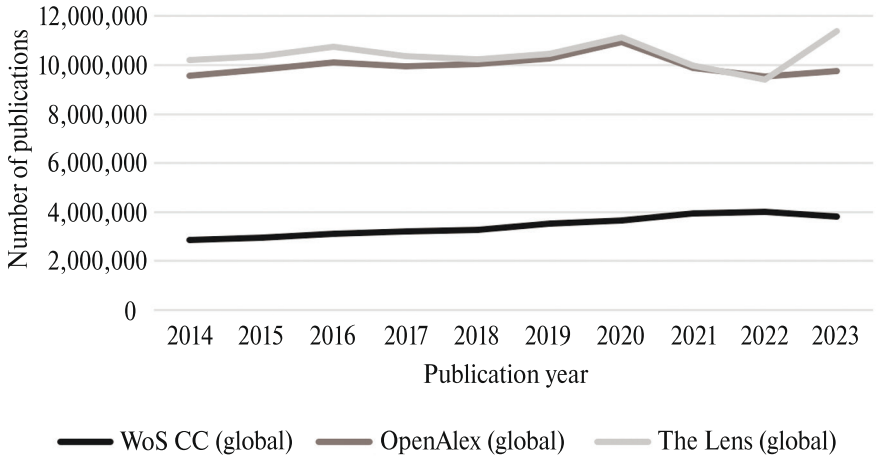

Fig. 1. Global flow of publications in OpenAlex and The Lens for 2014–2023 in comparison with WoS CC

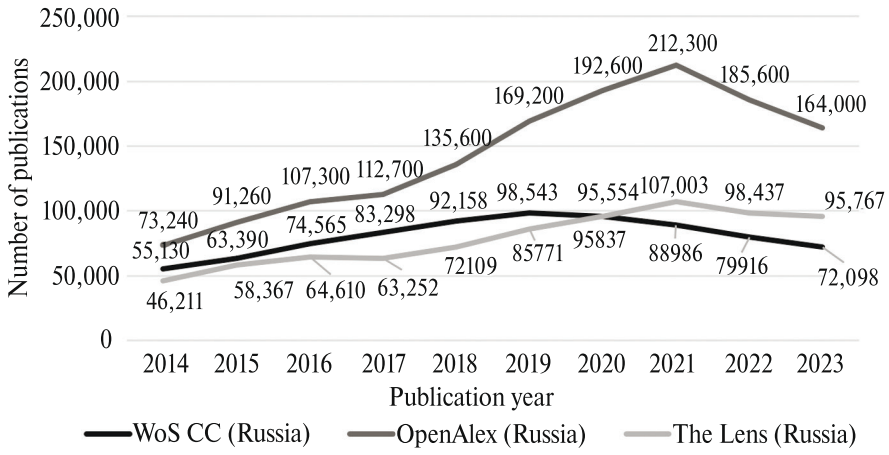

Fig. 2. The Russian flow of publications in OpenAlex and The Lens for 2014–2023 in comparison with WoS CC

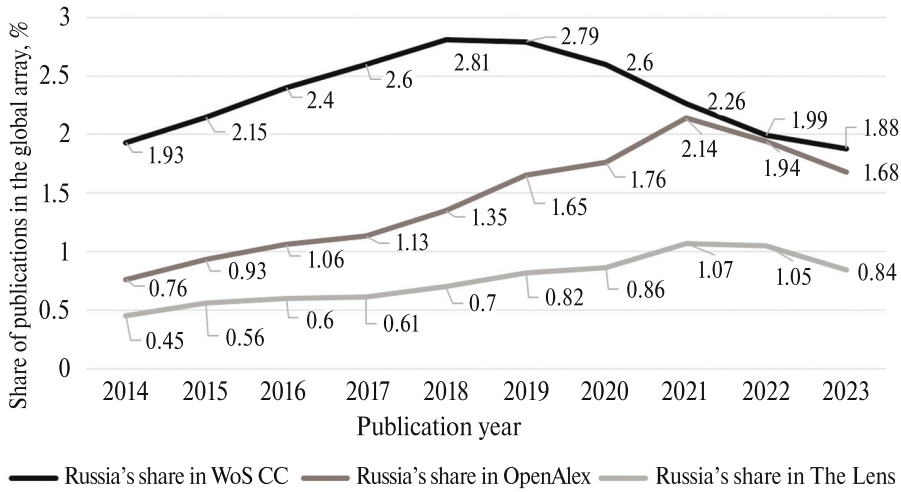

Fig. 3. Dynamics of the share of the Russian segment of publications in the global array on OpenAlex and The Lens in comparison with WoS CC in 2014–2023

from the most common classification at the journal level (WoS CC) to classifications at the level of individual publications (OpenAlex) [28]. Categories in different systems can be either more general or more specific, and the same name of child subject categories can appear in different parent categories. For example, in OpenAlex, the name of the child category "Biochemistry" appears in two parent categories – "Biochemistry, Genetics and Molecular Biology" and "Medicine". Despite the fact that the category names in OpenAlex and The Lens are mostly identical in overlap, statistics on the number of publications in overlapping subject categories vary from resource to resource: OpenAlex leads in some categories, and The Lens leads in others (Tables 3, 4). If the categories in these resources are mostly similar, then there are much fewer matches like this with WoS CC. Therefore, only the closest overlapping subject categories were selected to compare resources against the background of WoS CC.

Let us compare the Russian publication activity in various subject categories in OpenAlex and The Lens with each other and against the background of WoS CC.

OpenAlex against the background of WoS CC. When comparing the lists of subject categories in WoS CC and OpenAlex, 96 common categories were identified: 38% of the 252 child categories in OpenAlex and 37% of the 257 subject categories in WoS CC. It was found that in 60 overlapping categories, the average share of Russian publications for 2014–2023 in the global array in WoS CC was greater than in OpenAlex.

An analysis of the dynamics of the shares of Russian publications in overlapping subject categories in OpenAlex against the background of WoS CC showed that during the study period there was a decrease in the gap between the shares of Russian publications in the global array in OpenAlex relative to WoS CC. If in 2014–2018 WoS CC was the undisputed leader in terms of the share of the reflected Russian array against the global background, then in 2019–2023 the indicators of these two resources converged. However, there remains a significant number of scientific categories in which the share of the Russian array on the global background in WoS CC is significantly higher than in OpenAlex. Table 1 shows the subject categories in which the predominance of WoS CC over OpenAlex is particularly noticeable. The share of publications in 1.5% in WoS CC was chosen as the threshold value in order to cut off categories in which there was no noticeable publication activity.

As we can see, the most significant discrepancies between the shares of Russian publications in 2019–2023 in WoS CC and OpenAlex were noted in the categories "Information Science & Library Science" and "Forestry", by 17 and 14.5 times, respectively. The WoS CC database had an advantage over OpenAlex in reflecting Russian publications on natural sciences, engineering, and some humanities. The categories for which OpenAlex has the advantage are shown in Table 2.

So, in the "Cultural Studies" subject category, the share of Russian publications in OpenAlex outstrips WoS CC by more than 25 times, and in "Developmental and Educational Psychology" by 18 times. Most of the categories in which the share of Russian publications in 2019–2023 was higher in OpenAlex than in WoS CC were medical fields.

The Lens against the background of WoS CC. 63 overlapping subject categories were identified in The Lens and WoS CC, which corresponds to 39% of the 161 subject categories in The Lens and 24% of the 257 categories in WoS CC. In 61 such categories, the share of Russian publications in WoS CC was higher than in The Lens.

As shown in Fig. 3, the share of Russian publications in The Lens is the smallest compared to WoS CC and OpenAlex. Nevertheless, in this case, in 2019–2023, there was a slight reduction in the share gap between the Russian array in WoS CC and The Lens in almost all overlapping subject categories, although the alignment of indicators in the near future is unlikely to be achievable.

OpenAlex and The Lens: a comparison. OpenAlex and The Lens are open polythematic resources in which 133 overlapping subject categories of the same name have been identified.

Low share values of Russian arrays in The Lens relative to WoS CC and OpenAlex (see Fig. 3) look rather strange, bearing in mind that the number of global publications in The Lens and OpenAlex is almost identical. The weak representation of the Russian document flow in The Lens may be caused by such reasons as errors in the system that cause metadata loss during the indexing of documents containing information about the country; the specifics of the generated content; and bias against the Russian segment of documents.

The study showed that in 106 overlapping subject categories in OpenAlex, the share of Russian publications on a global background was higher than in The Lens (Table 3). In 43 categories, the advantage of OpenAlex was threefold or more, in 17 categories twofold. The categories of "Agronomy and Crop Science" (exceeding by more than 22 times), "Food Science and General Materials Science" (by 12 times) are particularly distinguished. The Lens is ahead of OpenAlex in only 27 categories (Table 4).

Thus, only in a small range of subject categories in technical sciences and life sciences, the share of Russian publications in the global array of 2019–2023 was higher in The Lens than in OpenAlex. The Lens' leadership (see Table 4) does not look as convincing as in the case of OpenAlex (see Table 3): the maximum difference in shares in 14 categories in The Lens is one to two times higher than OpenAlex, and only in two cases three and four times.

* * *

The study found that OpenAlex has advantages over another open access resource, The Lens. Despite the high absolute indicators of global flow dynamics in both of these databases, Russian publications in The Lens are severely underrepresented. This was clearly demonstrated by the analysis of the share distribution of Russian publications against the global background and across most overlapping subject categories.

At present, WoS CC remains the leader among the three resources considered in terms of the proportion of Russian and global arrays of publications. Nevertheless, starting in 2019, there has been an active reduction in the gap in shares between the Russian and global arrays of publications in WoS CC, OpenAlex, and The Lens. Since 2021, data on the dynamics of the Russian array against the global background according to OpenAlex correlate with the WoS CC indicators, which allows us to hope for high relevance of search results through OpenAlex.

REFERENCES

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. *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. https://doi.org/10.19181/smtp.2023.5.3.9 (In Russ.)
6. *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. https://doi.org/10.19181/smtp.2023.5.2.4 (In Russ.)
7. 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. *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. *Maltseva D.V., Pavlova I.A., Kapustina L.V.* 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.)
14. *Mazov N.A., Gureyev V.N.* Open Access Bibliographic Resources for Maintaining a Bibliographic Database of Research Organization // Scientific and Technical Information Processing. 2023, Vol. 50, No. 3, pp. 211–223. https://doi.org/10.3103/s0147688223030115 (In Russ.)
15. *Gureyev V.N., Ilicheva 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.)
16. *Tay A., Martín-Martín A., Hug S.E.* Goodbye, Microsoft Academic – hello, open research infrastructure? // Impact of Social Sciences Blog. May 27, 2021. http://eprints.lse.ac.uk/id/eprint/111325
17. 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
18. *Chawla D.S.* Microsoft Academic Graph is being discontinued. What's next? // Nature. 2021, June 15 2021.
19. *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
20. *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
21. *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. January 24.
https://www.nature.com/articles/d41586-022-00138-y
22. OpenAlex. Technical documentation.
https://docs.openalex.org/api-entities/works
23. *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/
24. CWTS Leiden Ranking Open Edition. https://open.leidenranking.com/
25. *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
26. *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. November 25, 2019.
https://doi.org/10.31229/osf.io/t56yh
27. The Lens:
https://en.wikipedia.org/wiki/The_Lens
28. OpenAlex: End-to-End Process for Topic Classification.
https://docs.google.com/document/d/1bDopkhuGieQ4F8gGNj7sEc8WSE8mvLZS/edit#heading=h.5w2tb5fcg77r