Metadata fields in Web of Science, Scopus, Dimensions and OpenAlex Databases: An exploratory analysis of the possibilities and ease of doing scientometric analyses

Prashasti Singh[1], Vivek Kumar Singh[2, 3, 4], Anurag Kanaujia[4]

[1]Shri Ram College of Commerce, University of Delhi, Delhi, India

[2]Department of Computer Science, Banaras Hindu University, Varanasi, India

[3]Department of Computer Science, University of Delhi, Delhi-110007, India

[4]Delhi School of Analytics, University of Delhi, Delhi-110007, India

ABSTRACT: *The various multidisciplinary and subject-specific scholarly databases have been the prime source of data for information retrieval and STI-related studies. Traditionally, Web of Science and Scopus have been popular choices of researchers. However, during the last few years, several new scholarly databases have been created, such as Dimensions, OpenAlex etc. Several studies have shown that coverage and data fields provided by these databases decide, to a large extent, the kind and scale of analysis that can be performed using these databases. However, there is a need for an updated study that includes the newer databases and focuses more on the availability of different metadata fields. This work is a step ahead as it provides a detailed comparative analysis of the research metadata fields provided by Web of Science, Scopus, Dimensions and OpenAlex databases. The analysis helps understand the possibilities and ease of scientometric analysis using these databases. Valuable results are obtained, and practical suggestions are provided for scientometric researchers, practitioners, and database managers.*

Keywords: Research Evaluation, Scientometric analysis, Scholarly Databases

Received: 31 August 2024, Revised 5 September 2024, Accepted 10 September 2024

DOI: https://doi.org/10.6025/stm/2024/5/214-222

1. Introduction

The most recent decade has recorded massive growth in scientific research which has led to the advent of many scholarly databases to index the research outputs. There are various applications of scholarly databases including information search and retrieval to research assessment at different levels. Traditionally, Web of Science and Scopus have been widely used by researchers for drawing data for various research evaluation purposes. However, during the last few years, newer academic databases like Dimensions, OpenAlex have emerged and were able to attract attention from the research community and practitioners. Research assessment exercises in Scientometric involve computation of different metrics namely, article-based metrics, author-level metrics, journal-level metrics, bibliographic/citation network creation, publication-citation patterns, open accessibility of research, collaboration at domestic/international level etc. The computation of these metrics requires obtaining metadata for publication indexed in scholarly databases.

The Web of Science is among the oldest scholarly databases having its origin from the citation index with its foundation dating back to 1955 at the Institute of Scientific Information (ISI) where Eugene Garfield created SCI which evolved into WoS in 1990s. It is currently owned by Clarivate Analytics and classifies scientific research across 254 subject disciplines in three major citation indices namely, SCIE (Science Citation Index Expanded), A&HCI (Arts & Humanities Citation Index) and SSCI (Social Sciences Citation Index). Scopus, a product by Elsevier was launched in 2004 and covers scientific journals, conferences, books etc. selected through a peer review process. A new database, Dimensions, was launched in 2018 by Digital Science, and studies have shown that it has fairly large coverage, not only of research publications but also of grants, policy documents, patents, datasets, etc. Another database, OpenAlex was launched in 2022, and is a successor of the Microsoft Academic Graph (MAG announced its retirement in 2021). Each of these databases provides a variety of metadata fields, which, to a large extent, decides what kind of scientometric analysis can be done by using these databases and with what ease.

The research metadata fields downloaded from these scholarly databases usually include dedicated fields for author details (for example, author name, researcher id, ORCID id, email address etc.), research organization details (for example, city, state & country details of the researcher/author's affiliating organization etc.), type of research publication (for example, journal paper, conference proceedings, erratum, article in press etc.), citation details (for example, the total number of citations accrued, cited references, different citation ratios etc.), details of funding organization (for example, funding text, grant number/id etc.), subject classification details (for example, research area and subject area of the publication) etc. The other commonly used metadata fields include doi, year of publication, language of publication etc. Though there have been studies comparing the coverage of various databases (Singh et al., 2021a) and the impact of database selection on country-level studies (Singh et al., 2021b), a comparative analysis of metadata fields in these databases is relatively less explored. Only some recent studies investigated metadata analysis, including the completeness and reliability of metadata fields across certain databases (Delgado & Ortega, 2024).

Therefore, there is a need for a comparative analysis of metadata fields provided by the major scholarly databases and to assess the impact of the availability of the different metadata fields on possibilities and ease of doing scientometric analysis. Recently, a study (Singh & Singh, 2023) did some initial work on comparing metadata fields of Web of Science, Scopus and Dimensions databases. This study attempts to extend and expand this work by providing a comprehensive comparative analysis of metadata fields of the four major scholarly databases (Web of Science, Scopus, Dimensions and OpenAlex. Among the major access routes provided by these databases, we have considered User Interface (UI) access routes, owing to their popularity across various geographical regions). The current study differs from the earlier study in terms of the inclusion of one of the new databases- OpenAlex. A comparative analysis of the metadata fields of all four databases is done to understand the impact of the metadata fields on possibilities and ease of doing scientometric analysis. Such an analysis can also provide useful insights into the organization of these databases.

2. Related Work

There have been some previous studies that attempted to compare different aspects of scholarly databases. Such aspects of comparison include journal coverage, citation coverage across databases, and comparison of other important features of databases, specifically the newer ones. For example, coverage analysis and overlaps between Web of Science and Scopus for university domain were performed in the study by Vieira & Gomes, (2009) while the disciplinary comparison of the coverage of Web of Science and Scopus was performed in some studies, such as Mongeon & Paul-Hus (2016), Gavel & Iselid (2008), Chadegani et al. (2013), Aksnes & Siversten (2019). A stable growth of publications and citations among Web of Science, Scopus and Google Scholar on a defined dataset for five broad disciplinary areas was concluded in the study by Harzing & Alakangas (2016). A recent study by Singh et al. (2021a) performed a comprehensive analysis of the journal coverage of Web of Science, Scopus and Dimensions databases and found that while Web of Science continues to remain selective, Dimensions has a much wider coverage of journals. Another study by Singh et al., (2022) examined the relationship between the number of journals indexed from a country in Scopus and the productivity ranking of those countries in Scopus.

The citations to articles covered by different databases were also analysed in some previous studies like Yang & Meho (2006) and Falagas et al. (2008). Citation studies on macro and micro-level analysis of South African journals in Web of Science, Scopus and Google Scholar were performed by Adriaanse & Rensleigh (2011, 2013). The studies by Martin-Martin et al. (2018a, 2018b) also analysed the citation counts across disciplines among Google Scholar, Web of Science and Scopus. A comparison of h-indices of highly-cited Israeli researchers and citation counts across Web of Science, Scopus and Google Scholar was performed in the study Bar-Ilan (2008).

Some recent studies have focused their attention on comparing newer databases with older widely used databases. The study by Thelwall (2018) was the first work to explore the Dimensions database, and it concluded that Dimensions could stand as a strong contender to the popularly used database Scopus for citation analyses. Dimensions as a potential rival to traditional academic databases were also noticed in the study by Visser, Eck and Waltman (2019), which performed a comprehensive comparative analysis of entire publication records of Web of Science, Scopus, Dimensions and Crossref. Exploration of differences among Web of Science, Scopus and Microsoft Academic at institutional level was performed in another study by Huang et al. (2020). It was found that the choice of just a single database could hamper institutional-level assessment and that Microsoft Academic had greater coverage than Web of Science and Scopus. Large-scale comparison of multiple multidisciplinary data sources has been performed by many more studies, such as Harzing (2019), Martin-Martin et al. (2021) and Visser, Eck and Waltman (2021).

For long, traditional databases like Scopus and Web of Science have remained at the helm for performing scientometrics analyses, typically evaluative bibliometrics. At times, these databases have been observed to under-represent many disciplines across the world (Archambault et al., 2006; Basson et al., 2022; Larivière et al., 2006). Recently, OpenAlex has got a lot of attention possibly because of its non-profit nature and open availability of source code. This has prompted several studies to prefer OpenAlex for bibliometric and altmetric studies (Akbaritabar et al., 2023; Haunschild & Bornmann, 2024) as well as for scoping reviews and for knowledge graph creation for AI innovation (Arroyo-Machado & Costas, 2023; Ismail et al., 2023; Mongeon et al., 2023; Massri et al., 2023). In terms of coverage, OpenAlex has a much broader coverage when compared to traditional databases (Scheidsteger & Haunschild, 2022; Scheidsteger et al., 2018). However, researchers who tried to assess the database from this perspective have noticed issues of data completeness and reliability in OpenAlex (Alperin et al., 2024)

Despite there being a good number of previous studies on different scholarly databases, their metadata fields have not been much explored in a comparative manner. Only one recent study (Singh & Singh, 2023) did some initial work on comparing metadata fields of Web of Science, Scopus and Dimensions databases. However, this study did not include the OpenAlex database in analysis. Therefore, the present study attempts to bridge this gap and provide a comprehensive comparative analysis of the metadata fields of the four major databases. Not only the structure and variation in the metadata fields are analysed, but their impact on possibilities and ease of doing scientometric analysis is also assessed.

3. Data & Method

In order to perform the analysis on research metadata fields, a few publication records were downloaded from each of the databases by formulating different queries in their UI interface. The publication data for the first 300 records was downloaded from each of the four databases, and from the downloaded records, the metadata fields were identified. Some examples of the metadata fields downloaded from the 4 databases include:

Web of Science: PT (Publication Type), AU (Authors), PY (Publication year of a research article), DOI (Digital Object Identifier for a publication), etc.

Scopus: Authors (Authors of the publication), Title (Title of the publication), Source title (the source in which the article is published, journal or other venue, etc.), Cited by (citations received by a publication record), etc.

Dimensions: rank (Ranking of a publication based on its relevance to the query), Publication ID (unique id allotted by Dimensions database to its indexed publication record), Open Access (whether an article is open access or not, it's open access status), etc.

OpenAlex: *author_ids, publication_year, referenced_works* etc.

The metadata fields thus obtained were then grouped into different categories like metadata fields for author details, publication details, citation details, research organization details, funding details, open access details, textual details etc. For these categories, the relevant metadata fields were identified from each database. The different metadata fields were then analysed from the point of view of different scientometric analysis requirements. Post this, the possibilities and ease of doing different scientometric analyses from the metadata fields obtained from the databases were identified. Figure 1 below briefly overviews the data and methodological steps used to perform the analysis.

4. Analytical Results

4.1. Identification and Characterization of specific metadata fields

The metadata fields that were obtained through a search in the UI interface of the four databases namely, Web of Science, Scopus, Dimensions and OpenAlex, were identified and then categorised into groups such as author related metadata, citation usage related metadata, publication details metadata, SDG related metadata, etc. The **figures 2, 3, 4** and **5** present the complete set of metadata fields provided through the UI download route of Web of Science, Scopus, Dimensions and OpenAlex.

The Web of Science database contained almost all the relevant metadata for article details except for the ranking of publications by relevance. For the other categories, too, relevant metadata was found; for example, metadata for usage count details (U1, U2) is a unique feature which was found only in the case of Web of Science. Similarly, the metadata fields for funding-

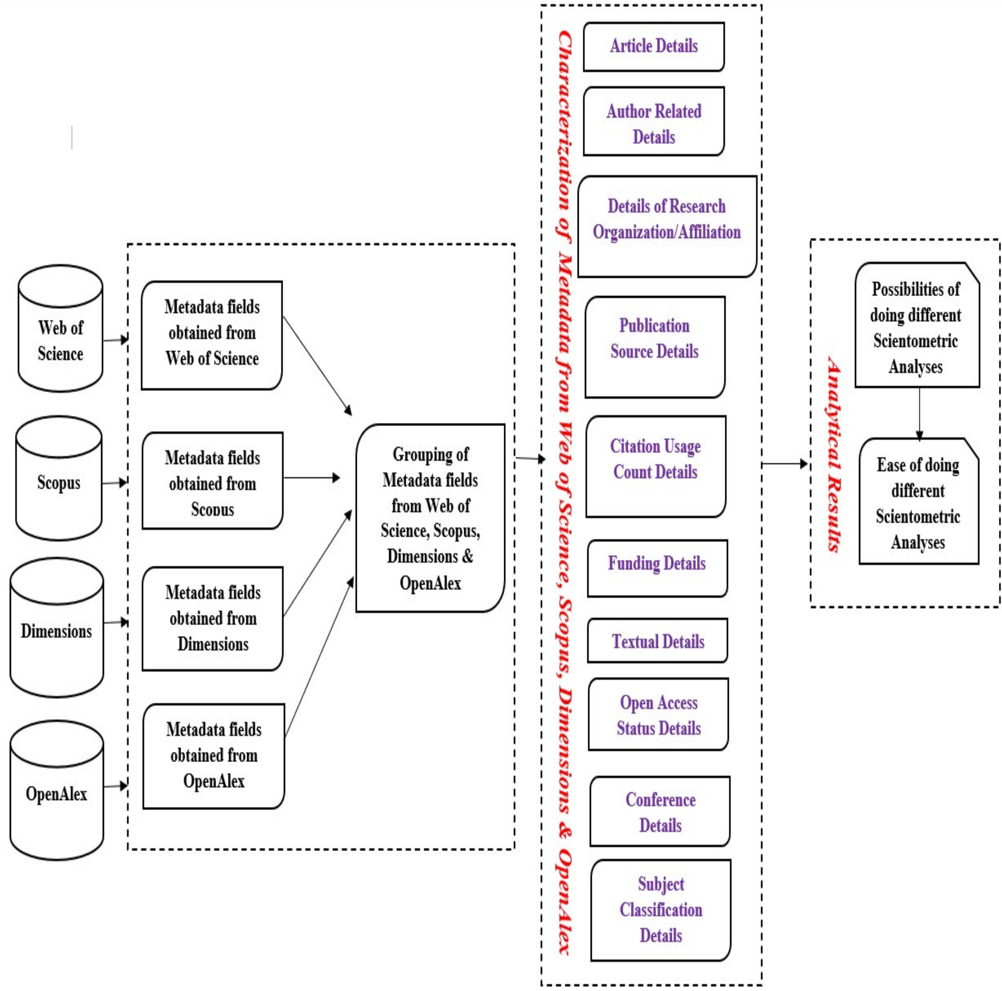

Figure 1. Diagrammatic overview of the methodology adopted

PT, AU, BA, BE, GP, AF, BF, CA, TI, SO, SE, BS, LA, DT, CT, CY, CL, SP, HO, DE, ID, AB, C1, C3, RP, EM, RI, OI, FU, FP, FX, CR, NR, TC, Z9, U1, U2, PU, PI, PA, SN, EI, BN, J9, JI, PD, PY, VL, IS, PN, SU, SI, MA, BP, EP, AR, DI, DL, D2, EA, PG, WC, WE, SC, GA, PM, OA, HC, HP, DA, UT

Figure 2. Metadata Fields obtained from Web of Science (about 71 fields)

Authors, Author full names, Author(s) ID, Title, Year, Source title, Volume, Issue, Art. No., Page start, Page end, Page count, Cited by, DOI, Link, Affiliations, Authors with affiliations, Abstract, Author Keywords, Index Keywords, Molecular Sequence Numbers, Chemicals/CAS, Tradenames, Manufacturers, Funding Details, Funding Texts, References, Correspondence Address, Editors, Publisher, Sponsors, Conference name, Conference date, Conference location, Conference code, ISSN, ISBN, CODEN, PubMed ID, Language of Original Document, Abbreviated Source Title, Document Type, Publication Stage, Open Access, Source, EID

Figure 3. Metadata Fields obtained from Scopus (about 46 fields)

Rank, Publication ID, DOI, PMID, PMCID, ISBN, Title, Abstract, Acknowledgements, Funding, Source title, Anthology title, Book editors, Publisher, ISSN, MeSH terms, Publication date, PubYear, Publication date (online), Publication date (print), Volume, Issue, Pagination, Open Access, Publication Type, Document Type, Authors, Authors (Raw Affiliation), Corresponding Authors, Authors Affiliations, Research Organizations - standardized, GRID IDs, City of standardized research organization, State of standardized research organization, Country of standardized research organization, Funder, Funder Group, Funder Country, 'UIDs of supporting grants, Supporting Grants, Times cited, Recent citations, RCR, FCR, Altmetric, Source Linkout, Dimensions URL, Fields of Research (ANZSRC 2020), RCDC Categories, HRCS HC Categories, HRCS RAC Categories, Cancer Types, CSO Categories, Units of Assessment, Sustainable Development Goals

Figure 4. Metadata Fields obtained from Dimensions (about 55 fields)

id, display_name, publication_date, relevance_score, primary_location_id, primary_location_display_name, primary_location_host_organization, primary_location_issns, primary_location_issn_l, primary_location_type, primary_location_landing_page_url, primary_location_pdf_url, primary_location_is_oa, primary_location_version, primary_location_license, author_ids, author_names, author_orcids, author_institution_ids, author_institution_names, is_oa, oa_status, oa_url, cited_by_count, doi, mag, pmid, pmcid, publication_year, cited_by_api_url, type, is_paratext, is_retracted, biblio_issue, biblio_first_page, biblio_volume, biblio_last_page, referenced_works, related_works, concept_ids, apc_paid_value, locations_license, primary_location_source_is_oa, primary_location_source_issn_l, apc_paid_currency, topics_score, primary_topic_display_name, authorships_author_position, locations_pdf_url, keywords_keyword, concepts_level, referenced_works_count, counts_by_year_year, fulltext_origin, primary_location_source_host_organization, best_oa_location_source_issn_l, primary_location_source_id, primary_topic_field_id, primary_topic_domain_display_name, best_oa_location_source_host_organization_name, locations_landing_page_url, sustainable_development_goals_score, apc_paid, primary_location_is_published, primary_topic_field_display_name, best_oa_location_source_display_name, language, grants_award_id, primary_location_is_accepted, mesh_is_major_topic, locations_count, best_oa_location_source_issn, counts_by_year_cited_by_count, grants_funder, mesh_qualifier_name, locations_is_oa, best_oa_location_source_is_in_doaj, mesh_qualifier_ui, best_oa_location_license, institutions_distinct_count, best_oa_location_is_oa, best_oa_location_is_published, mesh_descriptor_ui, best_oa_location_source_type, sustainable_development_goals_display_name, locations_version, primary_location_source_display_name, best_oa_location_source_host_organization, topics_display_name, best_oa_location_source_host_organization_lineage, cited_by_percentile_year_min, primary_location_source_host_organization_lineage_names, primary_topic_subfield_id, open_access_any_repository_has_fulltext, primary_location_source_type, primary_location_source_issn, primary_location_source, primary_location_source_host_organization_lineage, indexed_in, authorships_is_corresponding, concepts_display_name, abstract, type_crossref, apc_paid_provenance, apc_paid_value_usd, title, concepts_score, corresponding_author_ids, best_oa_location_is_accepted, apc_list_value_usd, best_oa_location_version, concepts_id, primary_location, best_oa_location_landing_page_url, locations_is_accepted, open_access_is_oa, primary_location_source_is_in_doaj, has_fulltext, keywords_score, sustainable_development_goals_id, topics_id, is_authors_truncated, primary_topic, cited_by_percentile_year_max, apc_list_currency, ids_doi, ngrams_url, corresponding_institution_ids, best_oa_location_source, apc_list, mesh_descriptor_name, best_oa_location_pdf_url, countries_distinct_count, primary_topic_score, ids_openalex, created_date, primary_topic_id, best_oa_location_source_id, primary_topic_subfield_display_name, locations_is_published, best_oa_location, apc_list_provenance, primary_location_source_host_organization_name, best_oa_location_source_host_organization_lineage_names, updated_date, grants_funder_display_name, authorships_raw_affiliation_string, apc_list_value, concepts_wikidata, locations_source, best_oa_location_source_is_oa, primary_topic_domain_id

Figure 5. Metadata Fields obtained from OpenAlex (about 152 fields)

related text, source of publication, basic fields for extracting citation information, basic details for conference publications etc. are also found in the metadata fields obtained from Web of Science In the Scopus database, the relevant metadata fields were found for article related details, author related details, funding acknowledgements, basic metadata containing citation details etc. While similar to Web of Science, Scopus also contained detailed metadata fields for conference information. However, in the case of citation details, it was found to have limited metadata fields as compared to the Web of Science. In the case of funding information, Scopus also provides explicit metadata for any supporting grants. Scopus metadata doesn't contain information on the subject classification of publication records, it could only be obtained through a lookup in the database for the particular record.

The Dimensions database also contains relevant metadata fields for categories such as article details, publication source details etc. Moreover, it has dedicated and structured metadata for research organization details wherein the city/state/country of the affiliating research organization is explicitly provided. Similarly, it has structured metadata for funding information as well. It contains basic metadata for citation details, RCR, FCR and altmetric details as well, but doesn't contain metadata for cited references. In order to obtain cited references for publication records, an extra lookup operation would be required from its UI interface through a download for bibliometric mapping. Also, language metadata and keywords for publication records are not present.

The publication metadata fields in OpenAlex are very diverse as compared to the other databases i.e. Web of Science, Scopus and Dimensions. Around 152 fields are provided that contain data classified under them. Apart from the fields that contain basic publication details, there are other fields that simplify the procedure of performing a scientometric analysis. For example, it contains the metadata which denotes a publication's SDG (Sustainable Development Goals) classification. It contains extra metadata fields for author details; for example, it contains metadata which contains details on the number of institutions involved in a publication (*institutions_distinct_count*), and it also contains details on the authorship position of authors involved in a publication (*authorships_author_position*). It has multiple metadata fields for publication source details as well as for open-access details. Interestingly, it contains detailed metadata on keywords and concepts of a publication record and the score related to the concepts extracted from publication records. Since, it is an open source and widely used in quantitative science studies and other related contexts, therefore there is a need for proper evaluation of specific metadata fields for their accuracy and reliability, their disambiguation and indexing data beyond journal articles. Data quality improvement is an important issue which needs to be addressed while using such a huge and diverse number of metadata fields provided by OpenAlex.

Appendix 1 provides an overview of the major metadata fields in the four databases, organized under different categories.

4.2. Possibilities and Ease of Doing Scientometric Analysis

The identification of the research metadata fields provided in the above table informs the major fields that contain information on author details, publication details, citation and usage details, funding details, etc. Now by an analysis of these fields from the perspective of scientometric analysis requirements, the possibilities of doing different types of scientometric analysis and the ease of doing so can be explored. **Table 1** below provides a comparative overview of the scientometric analysis possible from the metadata fields obtained from Web of Science, Scopus, Dimensions and OpenAlex. The symbol [*“*] indicates whether a specific scientometric analysis can be performed while [*X*] indicates that the specified scientometric analysis cannot be performed from the research metadata fields of the particular database. Similarly, [*E*] indicates that a specific scientometric analysis can be easily performed using the research metadata fields of the database, and [*M*] indicates that it can be done with a medium difficulty level.

5. Inferences

In view of the exploration of the metadata fields of the four major databases and the analysis of possibilities and ease of doing scientometric analysis, the following inferences can be drawn:

• Web of Science and Scopus continue to remain the most reliable and dependable databases in terms of data completeness and accuracy and, hence, remain the preferred choices for research evaluation exercises at different levels.

• Newer databases like Dimensions and OpenAlex have a lot of new features that provide better machine access and readability and are known to have higher coverage of research items. However, many studies have pointed out the issues of data completeness and reliability. Therefore, these newer databases may be interesting to explore for information retrieval purposes but are yet to mature to reach the level of being suitable for use in evaluative scientometrics, more so in the case of national and institutional contexts.

• In many cases, a newer database may have a very useful and new metadata field, but these fields are often not populated with data, making them useless for analysis.

Analysis topic/ question	Web of Science		Scopus		Dimensions		OpenAlex	
	Possible	Ease	Possible	Ease	Possible	Ease	Possible	Ease
Language Composition of Research Output	√	E	√	E	X	-	√	E
Retracted Paper Analysis	√	E	√	E	X	-	√	E
Exercises involving Subject Category Mapping	√	E	X	-	√	E	√	E
SDG Analysis	X	-	X	-	√	E	√	E
Analysis involving Article Usage	√	E	X	-	X	-	X	-
Text based Analysis using keywords	√	E	√	E	X	-	√	E
Analysis using conference publications	√	E	√	E	X	-	√	M
Analysis involving citation network creation and mapping (Co-citation, bibliographic coupling)	√	M	√	M	X	-	√	M
Gender based analysis of research publications	√	E	√	E	√	E	√	E
Collaboration Patterns (Domestic vs International)	√	M	√	M	√	E	√	M
UIG Collaboration Analysis	√	M	√	M	√	E	√	M
Analysis involving article size	√	E	√	E	√	E	√	E
Analysis involving references and cited reference count	√	M	√	M	X	-	√	M
Analysis involving Funding Details	√	E	√	E	√	E	√	E

Table 1. Comparing the analysis possibilities and ease in Web of Science, Scopus, Dimensions and OpenAlex

6. Conclusion

This study presented a comparative analysis of the metadata fields provided by the four major databases- Web of Science, Scopus, Dimensions and OpenAlex. The analysis helps identify the differences in metadata fields across these databases when they are accessed through the UI route. A summary of the possibilities and ease of doing different scientometric analyses using these metadata fields is provided**.** The findings provide helpful practical suggestions to database managers, scientometric researchers, etc. As a future work, one may consider a comparative analysis of the databases when they are accessed through other routes (such as API routes). The functionalities and capabilities of different databases could further vary through the API access route. Another future work possibility is the analysis of metadata completeness of the major databases, as the presence of a metadata field has no use unless it is populated with data.

References

[1] Adriaanse, L. S., & Rensleigh, C. (2011). Comparing Web of Science, Scopus and Google Scholar from an environmental sciences perspective. *South African journal of libraries and information science*, *77*(2), 169-178.

[2] Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. *The Electronic Library*, *31*(6), 727-744.

[3] Akbaritabar, A., Theile, T., & Zagheni, E. (2023). *Global flows and rates of international migration of scholars* (No. WP-2023-018). Max Planck Institute for Demographic Research, Rostock, Germany.

[4] Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, *4*(1), 1-21.

[5] Alperin, J. P., Portenoy, J., Demes, K., Larivière, V., & Haustein, S. (2024). An analysis of the suitability of OpenAlex for bibliometric analyses. *arXiv preprint arXiv:2404.17663*.

[6] Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329-342.

[7] Arroyo-Machado, W., & Costas, R. (2023, April). Do popular research topics attract the most social attention? A first proposal based on OpenAlex and Wikipedia. In *27th International Conference on* Science, Technology and Innovation Indicators (STI 2023). International Conference on Science, Technology and Innovation Indicators.

[8] Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, *74*, 257-271.

[9] Basson, I., Simard, M. A., Ouangré, Z. A., Sugimoto, C. R., & Larivière, V. (2022). The effect of data sources on the measurement of open access: A comparison of Dimensions and the Web of Science. *PLoS one*, *17*(3), e0265545.

[10] Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ebrahim, N. A. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *arXiv preprint arXiv:1305.0377*.

[11] Delgado-Quirós, L., & Ortega, J. L. (2024). Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies*, 1-36.

[12] Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, *22*(2), 338-342.

[13] Harzing, A. W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?. *Scientometrics*, *120*(1), 341-349.

[14] Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*, 787-804.

[15] Haunschild, R., & Bornmann, L. (2024). Usage of OpenAlex for creating meaningful global overlay maps of science on the individual and institutional levels. *arXiv preprint arXiv:2404.02732*.

[16] Huang, C. K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, *1*(2), 445-478.

[17] Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online information review*, *32*(1), 8-21.

[18] Ismail, S., Mermoud, A., Marechal, L., Orso, S., & David, D. P. (2023, April). Capturing Trends Using OpenAlex and Wikipedia Page Views as Science Indicators: The Case of Data Protection and Encryption Technologies. In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. International Conference on Science, Technology and Innovation Indicators.

[19] Larivière, V., Archambault, É., Gingras, Y., & Vignola Gagné, É. (2006). The place of serials in referencing practices:

Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*(8), 997-1004.

[20] Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 2021; 126(1): 871-906

[21] Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018a). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of informetrics*, *12*(4), 1160-1177.

[22] Martín-Martín, A., Orduna-Malea, E., & Delgado López-Cózar, E. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*, *116*(3), 2175-2188.

[23] Massri, M. B., Spahiu, B., Grobelnik, M., Alexiev, V., Palmonari, M., & Roman, D. (2023, April). Towards innograph: a knowledge graph for AI innovation. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 843-849).

[24] Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*, 213-228.

[25] Mongeon, P., Bowman, T. D., & Costas, R. (2023). An open data set of scholars on Twitter. *Quantitative Science Studies*, *4*(2), 314-324.

[26] Scheidsteger, T., & Haunschild, R. (2022). Comparison of metadata with relevance for bibliometrics between Microsoft Academic Graph and OpenAlex until 2020. *arXiv preprint arXiv:2206.14168*.

[27] Scheidsteger, T., Haunschild, R., Hug, S., & Bornmann, L. (2018). The concordance of field-normalized scores based on Web of Science and Microsoft Academic data: A case study in computer sciences. *arXiv preprint arXiv:1802.10141*.

[28] Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021a). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*, 5113-5142.

[29] Singh, P., Singh, V., K., Arora, P., Bhattacharya, S. (2021b). India's rank and global share in scientific research: how data sourced from different databases can produce varying outcomes? *Journal of Scientific & Industrial Research*, **80**(4), 336-346. 10.56042/jsir.v80i04.38273

[30] Singh, V. K., Singh, P., Uddin, A., Arora, P., & Bhattacharya, S. (2022). Exploring the relationship between journals indexed from a country and its research output: an empirical investigation. *Scientometrics*, *127*(6), 2933-2966.

[31] Singh, V.K. & Singh, P. (2023). Exploring the publication metadata fields in Web of Science, Scopus and Dimensions: Possibilities and ease of doing scientometric analysis. Proceedings of the 19th International Conference on Scientometrics and Informetrics, 2023; https://zenodo.org/records/11163664

[32] Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science?. *Journal of informetrics*, *12*(2), 430-435.

[33] Vieira, E. S., & Gomes, J. A. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, *81*, 587-600.

[34] Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative science studies*, *2*(1), 20-41.

[35] Visser, M. S., van Eck, N. J., & Waltman, L. (2019, September). Large-scale comparison of bibliographic data sources: Web of Science, Scopus, Dimensions, and Crossref. In *ISSI* (Vol. 5, pp. 2358-2369).

[36] Yang, K., & Meho, L. I. (2006). Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for information science and technology*, *43*(1), 1-15.