https://doi.org/10.51408/issi2025_068

Linking Research Publications to SDGs: Exploring the SDG Mapping in Web of Science, Scopus and OpenAlex

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

[1]*prashasti.singh8@gmail.com*
Department of Computer Science, Shri Ram College of Commerce, University of Delhi, Delhi-110007 (India)

[2]*vivekks12@gmail.com*
Department of Computer Science, University of Delhi, Delhi-110007 (India)
NITI Aayog, Government of India, Delhi-10007 (India)

[3]*anuragkanaujia01@gmail.com*
Delhi School of Analytics, University of Delhi, Delhi-110007 (India)

[4]*abhirupnandy.online@gmail.com*
Institute of Informatics and Communication, University of Delhi, Delhi-110067 (India)

Abstract

The Sustainable Development Goals (SDGs) are 17 global objectives proposed by the United Nations to create a better and more sustainable future by 2030. Since the adoption of SDGs in 2015, several missions and programmes have been initiated across different countries towards achieving the relevant targets under SDGs. The advancements in science and technology research and development is believed to play a crucial role in achieving these targets. Motivated by the role of scientific outcomes in SDGs, various scholarly databases have started to map the indexed research publications to one or more of the SDGs. A host of approaches (employing keyword based, machine learning, manual curation etc.) have been used to link and map research publications under the SDGs. Some initial studies have shown that the mapping of publications in SDGs vary significantly across different databases. However, the classification accuracy, thematic focus, practical applicability, and impact of these classification approaches have not been studied well. Therefore, this work attempts to make a deeper exploration of the SDG mapping in three major scholarly databases- Web of Science, Scopus and OpenAlex and provide useful insights. For this purpose, a large-scale data sample of publications for the year 2023 obtained from these three databases are analysed on different aspects. Results suggest that not only the three databases vary significantly in terms of their individual SDG mapping, but there are also significant differences in the SDG-wise distribution and interlinkages across different SDGs. A divergence score to measure divergence of classification across the three databases is defined and computed. Finally, the probable reasons and factors that may be resulting in the variations in SDG mappings across the three databases are explored and discussed.

Introduction

The Sustainable Development Goals (SDGs) are 17 global objectives established by the United Nations in 2015, aiming to create a better and more sustainable future by 2030. These goals address critical challenges like no poverty (SDG 01), zero hunger (SDG 02), good health & well-being (SDG 03), quality education (SDG 04), gender equality (SDG 05), clean water and sanitation (SDG 06), affordable and clean energy (SDG 07), decent work and economic growth (SDG 08), industry, innovation and infrastructure (SDG 09), reduced inequalities (SDG 10), sustainable cities and communities (SDG 11), responsible consumption and production (SDG 12), climate action (SDG 13), life below water (SDG 14), life on land (SDG 15), peace, justice and strong institutions (SDG 16), and partnerships for the goals (SDG 17). Since their adoption in 2015, they have made a universal call to action and motivated significant research and development of novel technologies (United Nations, 2015). Some of the targets are interlinked and require global cooperation and partnership to achieve, promoting inclusive and sustainable development for all (Sachs, 2012).

The advancements in science and technology research and development are believed to play a crucial role in achieving several targets under SDGs (IISD, 2021; Singh et al., 2024). Research provides evidence-based insights, fosters innovation, and enables the development of sustainable and scalable interventions. It also promotes collaboration among governments, academia, the private sector, and civil society. The role of technology in achievement of SDGs has been studied and underlined (IISD, 2021). Motivated by the role of scientific outcomes in SDGs, various scholarly databases have started to map the indexed research publications to one or more of the SDGs. Scholarly databases are the primary source of providing metadata of scientific outcomes (such as publications and patents), to help identify outcomes that are related to SDGs. A host of approaches (employing keyword based, machine learning, manual curation etc.) have been used to link and map research publications under different SDGs.

Some initial studies have shown that the mapping of publications in SDGs vary significantly across different databases. For example, Armitage, Lorenz, Mikki, (2020) explored the SDG mapping of scholarly publications to identify whether independent bibliometric approaches yield the same results. Another study (Purnell, 2022) compared different methods of identifying publications related to the Sustainable Development Goal 13 on Climate Action and found significant variations across them. Some other studies (such as Hajikhani, & Suominen, 2022; Kashnitsky et al., 2024) also tried to explore SDG mapping of STI outputs in various respects. However, the classification accuracy, thematic focus, interlinkages, and divergence of classification across different scholarly databases have not been studied well. Therefore, this work attempts to make a deeper exploration of the SDG mapping in three major scholarly databases- Web of Science, Scopus and OpenAlex and provide useful insights. This study uses a large-scale data sample of publications for the year 2023 obtained from these three databases and attempts to characterize the SDG mappings in the three databases, not only in terms of variations but also in respect of interlinkages and classification divergence.

Related Work

The research addressing challenges of sustainable development is fairly distributed across the different subject areas. As the impact of challenges covered under the SDGs (e.g., poverty, healthcare, water & sanitation, gender equality and climate change) is becoming more obvious, more research attempting to address the related challenges has been undertaken across the world (Moyer & Hedden, 2020). However, like any other collective exercise, in absence of proper tracing and recapitulation, the cumulative impact from these exercises may end up being a zero-sum game. In academic research, the estimations have focused on the engagement of researchers and academic staff in universities across the world as key players in promoting SDGs. Studies have provided useful methodologies and insights for analytical exercises in assessment of research in MDGs and SDGs over the years. Keyword analysis is a widely used method for research topics and knowledge mapping. This has been utilized by authors to identify and study the research in SDGs (Armitage, Lorenz, & Mikki, 2020 and Bautista-Puig *et al.*, 2020).

Studies on publication metadata for global research output have explored major trends in research publications. Between the period of 2015 to 2019, the United States, United Kingdom and China are among the top three active countries for research in the different SDGs. SDG 17 i.e., partnerships for SDGs, has the most research publications associated with it, followed by SDG 13, i.e., climate action. Other SDGs with high research activity include SDG 12 (responsible consumption and production), SDG 15 (life on land), SDG 3 (good health and well-being), and SDG 1 (no poverty) (Sweileh, 2020). Co-citation occurrences showed that SDG 11 and SDG 3 are closely related and are frequently referred together in research publications (Bautista-Puig *et al.*, 2020). Most of the research corresponding to SDGs is published from research areas of Life sciences & Biomedicine, and Social Sciences (Meschede, 2020). These studies have started the important process of exploration of research trends in SDGs, and cover only a bird's eye view of the global research trends.

The study by Armitage, Lorenz, Mikki, (2020) explored the SDG mapping of scholarly publications to identify whether independent bibliometric approaches get the same results. Purnell (2022) compared four methods of identifying research publications related to United Nations Sustainable Development 13. These and some other studies identified variations across different scholarly databases. However, the classification accuracy, thematic focus, SDG classification interlinkages, and divergence of classification across major scholarly databases is yet to be suitably explored and analysed. This work attempts to address some of these gaps and provide useful insights about the SDG mapping of research publications across the three major databases.

Data & Method

The study utilized a large-sized data sample of research publications obtained from the three databases- Web of Science (WoS), Scopus, and OpenAlex through their user interfaces (UI). It focused on publication records from the country "India" for the year 2023, limited to document types "article" and "review,". India is one of the major advocates of Sustainable development and has significant focus on research in SDGs (Singh *et al.*, 2022). Further, the practical considerations of availability of data access has also motivated us to use this data as a sample for analysis. The data related to each of the SDGs was downloaded from all the three databases using appropriate search queries or filters provided by the databases. The choice of databases was motivated by the facts that Scopus and Web of Science are among the most popular and reliable databases for scholarly data. Open Alex, has emerged as a large repository of scholarly data providing openly accessible data for Scientometric application and other research purposes. In addition, these three databases provide SDG classification of records which was an essential consideration. While WoS and OpenAlex offer SDG-based filters in their UI, Scopus provides pre-formulated queries[1] for each SDG enabling data downloads with additional filters. The downloaded search results were pre-processed to eliminate NaN values and duplicate DOIs across all three databases. After this step, the unique records mapped with an SDG were 79,099 in WoS, 69,834 in Scopus and 214,873 in OpenAlex. The processed records in each database were then analysed to understand different patterns. The process of computing analytical results is detailed below.

The SDG wise distribution of publication records in each database is determined. Thereafter, the overlapping and unique DOIs for all SDGs across the three databases was identified. The SDG wise mapping of common DOIs across each pair of databases is computed, i.e. for WoS-Scopus, Scopus-OpenAlex and WoS-OpenAlex. The next step was to identify the interlinkages of classification. For this purpose, for a given database, each publication record was scanned to see in which other SDGs it is classified. In this way, all the interlinkages of SDG classification for publications from a given database were done. Similar process was followed for the other databases. The linkages were plotted on a network diagram. The edge weights are proportional to the number of publication records classified in the two connected SDGs together. Absence of any edge in a graph indicates that the two SDGs have no commonly tagged records.

The next computation involved calculating the divergence in SDG mapping of the three databases. Here, first the proportionate share of publication records classified in only one SDG, or two SDGs, or three, or more, is determined for all the three databases. Thereafter a numerical value of Divergence score is proposed and computed. The score can be defined as follows: Given a database, having x number of records mapped to a given SDG, the sum total of all other SDGs in which these x publication records are mapped is used to calculate a Divergence value for the given SDG. Similar values are computed for all other SDGs. Then, all such values are aggregated to compute the overall divergence score of that database. The divergence score for other databases can be computed in a similar way.

$$\square\square\square\square\square\square\square\square\square\square\square\ \square\square\square\square\square\ (\square\square\square\square) = \frac{\sum_{\square=1}^{17} \square_{\square\to\square}}{\square\square_{\square\square\square\square}} \quad \ldots(1)$$

$$\square\square\square\square\square\square\square\square\square\square\square\ \square\square\square\square\square = \frac{\sum_{\square=1}^{17} \square\square\square\square}{\square\square.\square\square\ \square\square\square\square} \quad \ldots(2)$$

where i = SDG number for which divergence is to be calculated, and $x_{i\to j}$ = number of records tagged under SDGi and SDGj, and, TP_{SDGi} = total number of records in the particular SDGi in the given database).

[1] Elsevier SDG Mapping: https://elsevier.digitalcommonsdata.com/datasets/y2zyy9vwzy/1

Results

Variation in Publication Volume mapped with SDGs across the three databases

In terms of total publication count; WoS reported 124,266 research publications for India during 2023, Scopus 194,965 and OpenAlex 340,900 for article and review types. After pre-preprocessing the SDG-wise downloaded data for duplicate and NaN DOIs; 79,099 DOIs were found for WoS, 69,834 for Scopus and 214,873 for OpenAlex. Thus, WoS and OpenAlex were seen to comprise a comparable share of SDG publications in total publications of India amounting to approximately 63% while Scopus reported 35.82% of SDG publications in total publications for India (**Table 1**).

Table 1. Total Publications and Publications tagged with a SDG.

Database	TP	Publications tagged with a SDG	%age
WoS	1,24,266	79,099	63.65
Scopus	1,94,965	69,834	35.82
OpenAlex	3,40,900	2,14,873	63.03

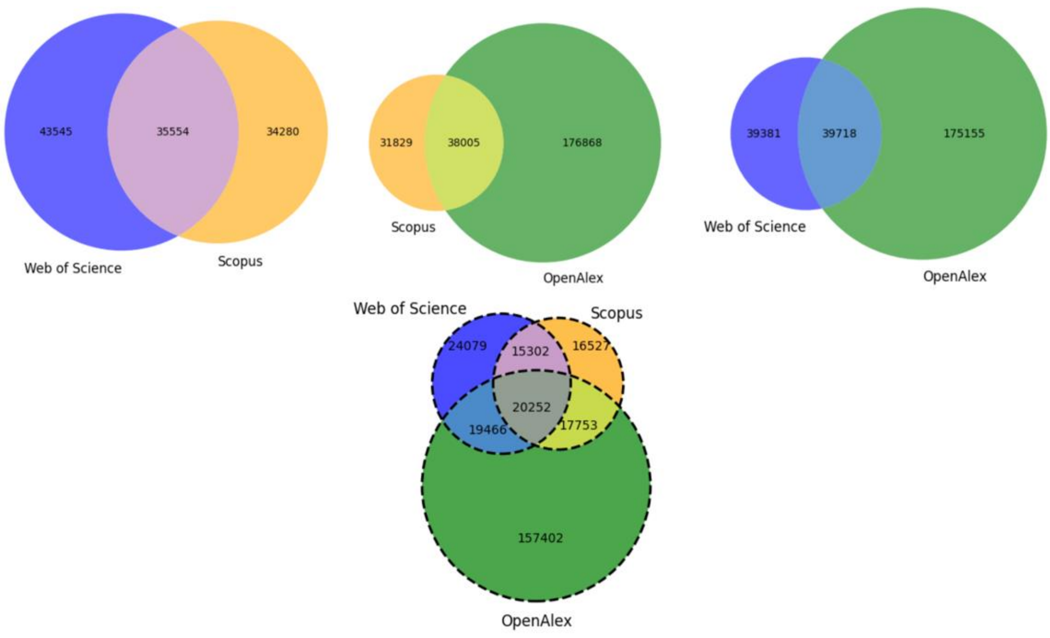


Figure 1. Overlaps in SDG tagged data across the three databases.

Thereafter, the pairwise overlaps across the databases as well as overall overlap across all the three databases together was also computed (**Figure 1**). The relevant numbers can be summarised as follows:

Web of Science- total pre-processed DOIs = 79,099

- Overlap with Scopus- 35,554 (44.95% of WoS), 43,545 DOIs are non-overlapping (55.05% of WoS)

- Overlap with OpenAlex- 39,718 (50.21% of WoS), 39,381 DOIs are non-overlapping (49.79% of WoS)
- Unique DOIs in WoS- 24,079 (30.44% of WoS)

Scopus- total pre-processed DOIs = 69,834

- Overlap with WoS- 35,554 (50.91% of Scopus), 34,280 DOIs are non-overlapping (49.09% of Scopus)
- Overlap with OpenAlex- 38,005 (54.42% of Scopus), 31,829 are non-overlapping (45.58% of Scopus)
- Unique DOIs in Scopus- 16,527 (23.67% of Scopus)

OpenAlex- total pre-processed DOIs = 2,14,873

- Overlap with WoS- 39,718 (18.48% of OpenAlex), 1,75,155 are non-overlapping (81.52% of OpenAlex)
- Overlap with Scopus- 38,005 (17.69% of OpenAlex), 1,76,868 are non-overlapping (82.31% of OpenAlex)
- Unique DOIs in OpenAlex- 1,57,402 (73.25% of OpenAlex)

Distribution of Publications across different SDGs in the three databases

The publication records classified under one or more SDGs were analysed for the distribution across SDGs. The proportionate share of publication records classified in each SDG in all the three databases was computed (**Figure 2**). Across the three databases, SDG 03 has the highest number of records (WoS: 45,316, Scopus: 31,587 and OpenAlex: 49,951) and SDG 07 has the second most number of records (WoS: 10,539, Scopus: 13,881 and OpenAlex: 33,855). In the third position however, WoS has SDG 13 with 9,703 records, Scopus has SDG 09 with 8,326 records and Open Alex has SDG 02 with 21,830 records. SDGs 01, 04, 08, 10, 16 and 17 have less than 1000 records classified in WoS which is less than 1% of total records.

A bird’s eye view picture of publication records shows differences in the proportional share of records mapped to the SDGs among the three databases **(Figure 2)**. Major variations are seen in case of WoS where more than half of the mapped records have been associated with SDG 3 (Good Health and Well Being) followed by SDG 07 (13.3%) and SDG 13 (12.2%); while SDG 4 (Quality Education), SDG 10 (Reduced Inequalities), SDG 16 (Peace, Justice and Strong Institutions) and SDG 17 (Partnerships for the Goals) have significantly lower number of associated records. In Scopus, SDG 3 has about 45.2% records mapped followed by SDG 7 (19.8%) and SDG 9 (11.9%). Remaining records are mapped mainly to SDG 2, 6, 8, 11, 12, and 13. On the other hand, in Open Alex, the highest percentage of records mapped to one SDG is 23.2% for SDG 3, followed by SDG 7 with 15.7%. This variation in the mapping of records in the databases may be an indication that in addition to their coverage variations, the schemes they utilize for SDG mapping are also different.

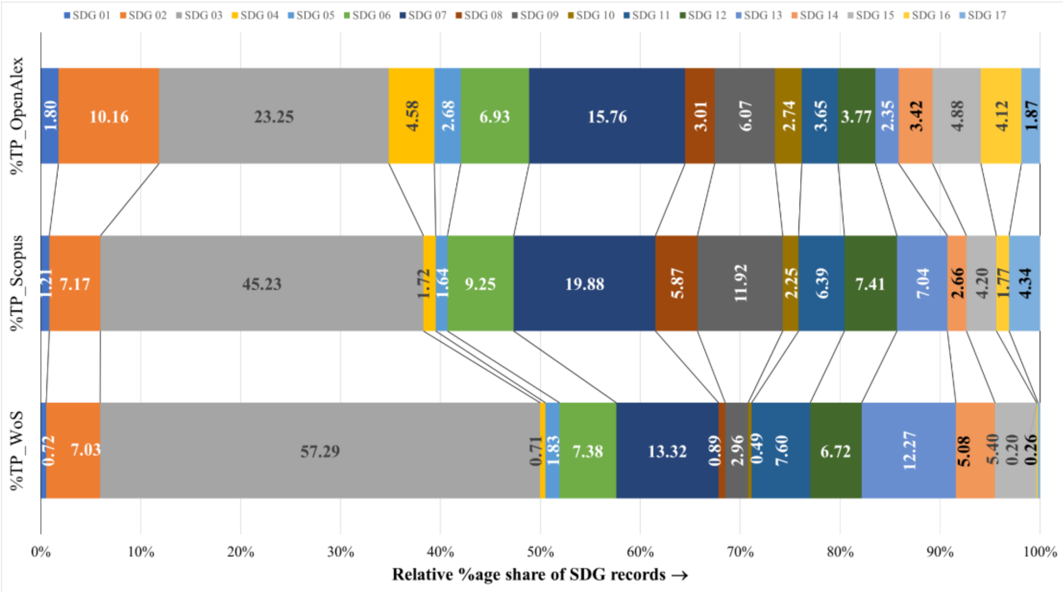
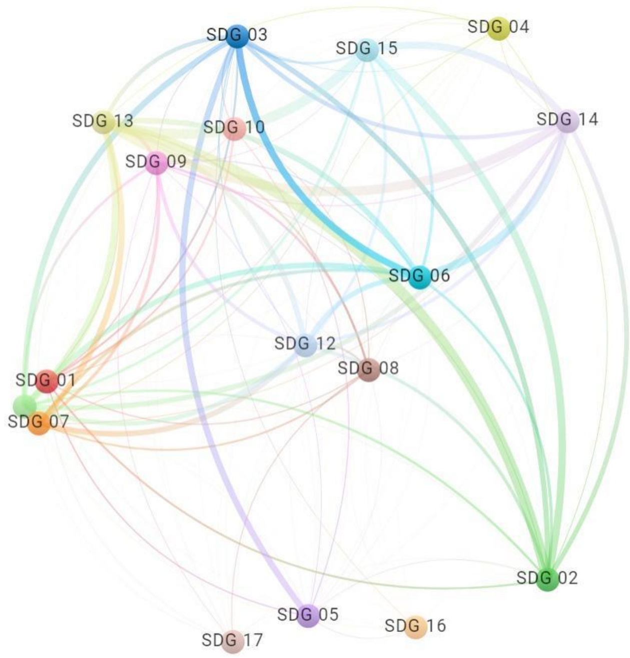


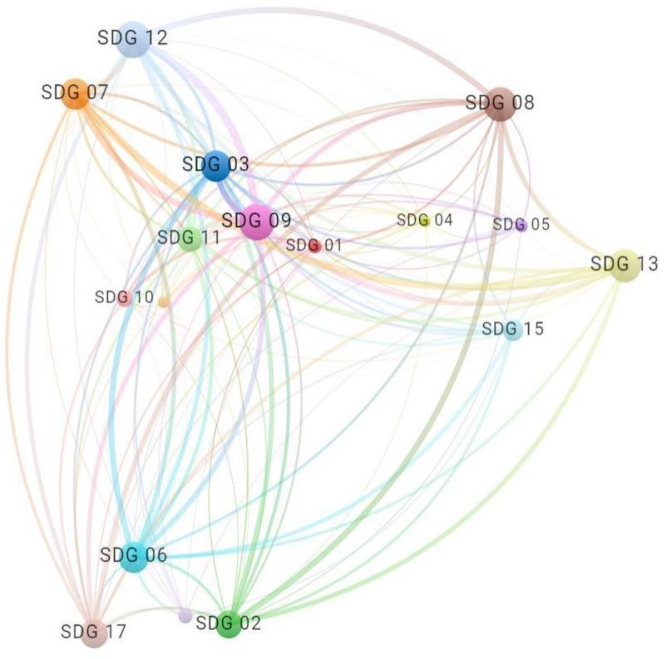
Figure 2. SDG wise tagging distribution of publications across three databases.

SDG mapping linkages across three databases

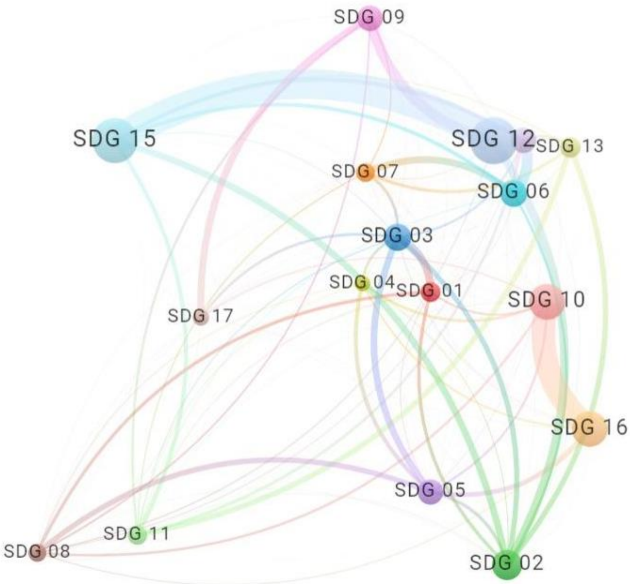
The mapping of records to each SDG in the selected database was plotted on a network map with vertices representing the total number of records assigned to the selected SDG and edges connecting two vertices showing the number of publication records which are mapped to both the SDGs **(Figure 3)**. As the edges in these maps visualise only the records mapped to multiple SDGs they provide a clearer and more detailed view into the mapping approaches of each database. In the case of WoS, edges from SDG 13 - 14 (edge weight, Wt. 2334), SDG 13 - 15 (Wt. 2542), SDG 13 - 02 (Wt. 2890), and SDG 13 - 11 (Wt. 1574), from SDG 14 - 15 (Wt. 1538) and, from SDG 15 - 02 (Wt. 1657) have the highest weights. In the case of Scopus, records are mapped together mostly between SDG 09 - 12 (Wt. 1822), SDG 09 - 07 (Wt. 1361), SDG 07 - 13 (Wt. 1663), and SDG 08 - 12 (Wt. 1293). While in Open Alex, edges between SDG 12 - 15 (Wt. 539), 09 (Wt. 122), SDG 10 - 16 (Wt. 382) and SDG 14 - 06 (Wt. 126) show a majority of records mapped together. Thus, a variation is also seen in SDG mapping linkages across the databases with common records mapped pairs among SDG 02, 11, 13, 14 and 15 in WoS; SDG 07, 08, 09, 12 and 13 in Scopus and, SDG 06, 09, 10, 12, 14, 15 in Open Alex.



(a) Web of Science



(b) Scopus



(c) Open Alex

Figure 3. SDG Interlinkages across three databases. (created using VoS Viewer)

Divergence of SDG mapping across the three databases

A further in-depth analysis of records in each database was conducted to find out records associated with multiple SDGs. This was done to analyze the focused or divergent nature of the mapping. Across the three databases, different relative proportions of records were observed to have mapping in multiple SDGs. The different pie charts plotted for each of WoS, Scopus and OpenAlex databases in **Figures 4a**, **4b** and **4c** provide further insights into SDG classification of DOIs in these databases. The pie charts depict the percentage share of DOIs classified into more than single SDG by a database.

In Open Alex, most of the records (98.92%) are mapped to only one SDG, in WOS and Scopus, 17.08% and 24.03% records are mapped to two or more SDGs. The results have been computed w.r.t. to 79,099 unique DOIs in WoS, 69,834 unique DOIs in Scopus and 214,873 unique DOIs in OpenAlex across all SDGs. From these charts, it can be seen that OpenAlex has the maximum percentage of DOIs tagged into a single SDG (98.93%) followed by WoS that classifies 82.92% of DOIs tagged into a single SDG and then Scopus which classifies approx. 76% of DOIs into a single SDG. It is observed that in WoS, the maximum limit to which the DOIs are tagged into more than single SDG is 9 while for Scopus it is 11 i.e. at most a DOI is tagged into 9 SDGs by WoS and 11 SDGs by Scopus. For OpenAlex, the maximum number of SDGs a DOI is tagged into other SDGs is 6 that too is a very minor percentage. Mostly, DOIs are classified into a single SDG by OpenAlex followed by DOIs classified into two SDGs while a very little percentage of DOIs are classified into three (3), four (4) and six (6) SDGs.

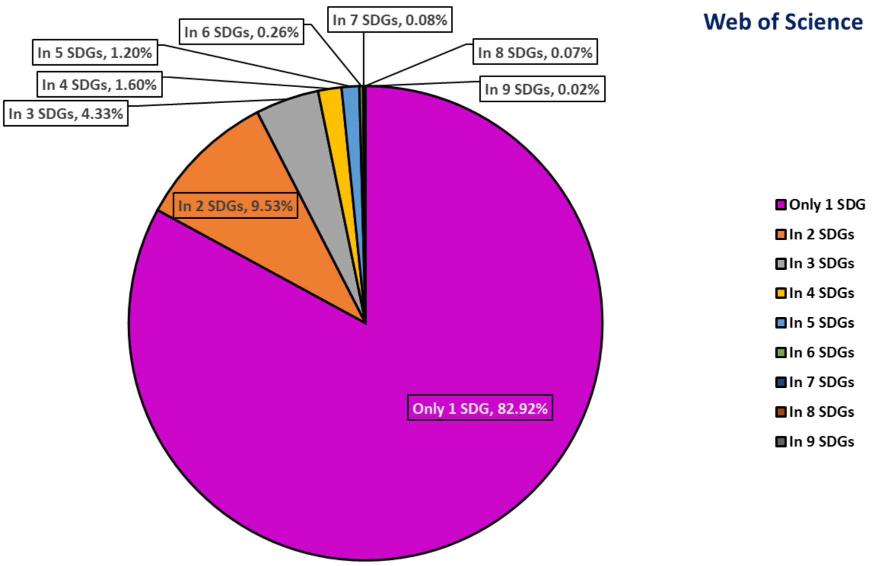


Figure 4a. Proportionate share of Publication Records mapped in single and multiple SDGs (Web of Science- 79,099 unique DOIs).

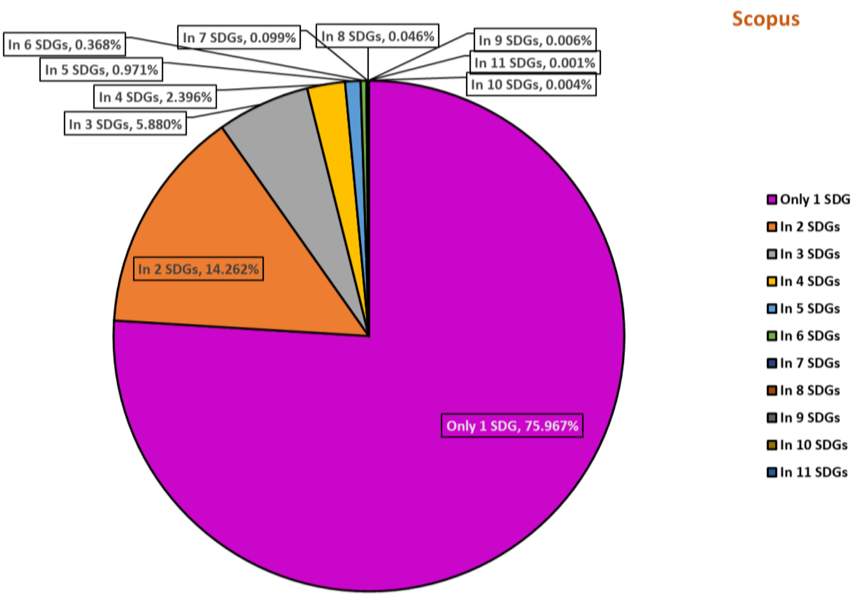


Figure 4b. Proportionate share of Publication Records mapped in single and multiple SDGs (Scopus-69,834 unique DOIs).

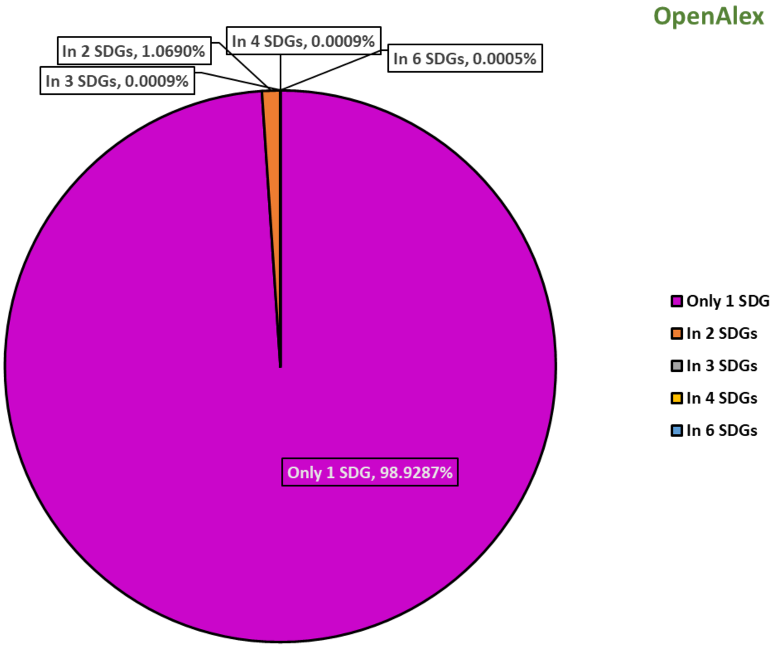


Figure 4c. Proportionate share of Publication Records mapped in single and multiple SDGs (OpenAlex- 214,873 unique DOIs).

The divergence of classification of publication records into different SDGs in each database has been analyzed next **(See Appendix)**. The methodology for computation of the divergence values for the 17 SDGs in each database is explained in the relevant section above. Results indicate that the divergence values for WoS range from 1.16 to 3.3, for Scopus it ranges from 1.23 to 3.2 while for OpenAlex it ranges from 1.00 to 1.09. Minimum value of divergence is observed for SDG 03 in both WoS and Scopus while OpenAlex has the minimum value of divergence value for SDG 07. However, the maximum divergence value is observed for SDG 14 in WoS, for SDG 01 in Scopus and for SDG 12 in OpenAlex. This indicates that a lower proportion of DOIs classified in SDG 03 in WoS and Scopus are classified into other SDGs and a lower proportion of DOIs classified in SDG 07 in OpenAlex are classified into other SDGs. Similarly, DOIs classified in SDG 14 in WoS are the ones to have been classified into more SDGs as compared to DOIs classified under other SDGs. DOIs classified in SDG 01 in Scopus are the ones to have also been classified into more SDGs as compared to DOIs classified under other SDGs. Higher proportion of DOIs classified in SDG 12 in OpenAlex have been classified into other SDGs. Also, WoS and Scopus have a mean divergence score of 2.36 and 2.38 respectively while OpenAlex has a mean divergence score of 1.03.
The range of scores and their depiction in the box plot (**Figure 5**) indicates that the SDG classification in WoS is more divergent than that in Scopus while the least divergent SDG classification is shown by OpenAlex. This implies that publication records classified under a particular SDG in OpenAlex are less likely to be mapped to other SDGs too. These differences can be attributed to the difference in the

schemes of SDG classification deployed in each database. This result further supplements the divergence of the databases in terms of SDG classification, where OpenAlex is found to be the least divergent one, followed by WoS, while Scopus is the most divergent in which more publication records are classified into multiple SDGs.

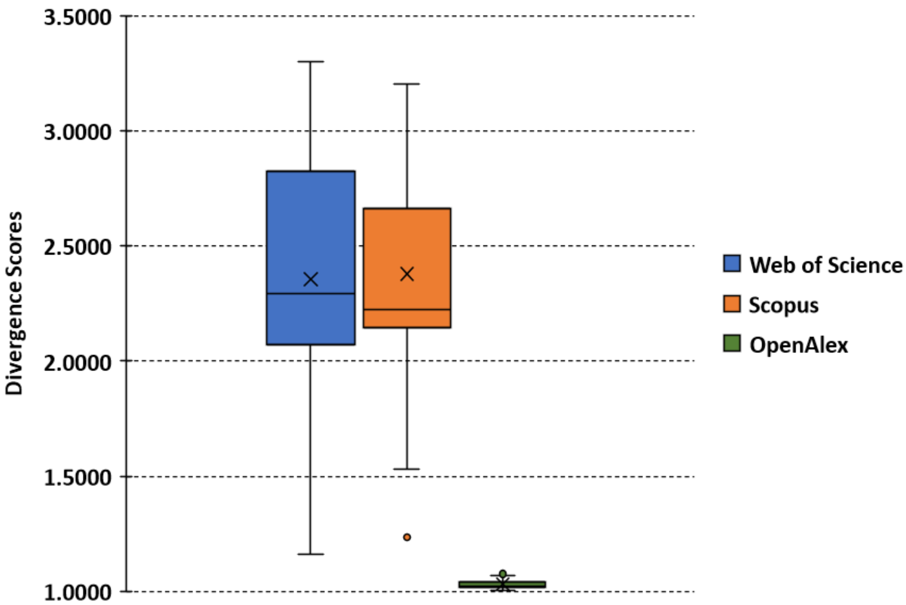


Figure 5. Range and Mean of Divergence Values for WoS, Scopus and OpenAlex.

Discussion

The study has analysed SDG mapping of publication records across the three databases, more specifically on the parameters of variations, interlinkages, and classification divergence. The results suggest variations in linking of research publication metadata to the 17 SDGs. Across the studied databases, the proportion of publications mapped with SDGs varies significantly indicating that the classification approaches used by them put varied attention to the different SDGs. *While 63% of records in Open Alex and WoS are mapped to SDGs, only 35% publication records in Scopus are associated with an SDG*. This could be a result of the keyword-based search approach used in Scopus (Bedard-Vallee *et al.*, 2023) which would be limited and slow to adapt to the continuously changing SDG research landscape. As a result, the coverage of mapped records in Scopus will only improve when the search queries are augmented. In contrast, Open Alex and WoS have the classification done at the backend level providing them flexibility to modify the classification criteria iteratively. This however places a restriction on the users of the database.

The three databases have a significant number of records which are not mapped to the same SDGs across them. An in-depth analysis of SDG mapping of records underlined the variations in database's approaches. In terms of individual SDGs, the highest number of records categorised under SDG 03 (Good Health and Wellbeing) across the databases but the proportionate contributions range from 23% in Open Alex to 57% in WoS. *In each database, the proportion of records mapped to individual SDGs varies significantly when compared to the other databases.* This variation may be due to the differences in mapping approaches used in these databases, namely keywords-based search query in Scopus, keyword based filters in WoS, and machine learning based in Open Alex. Upon exploring the keywords used in SDG classification approach for SDG 3 it was observed that Scopus Query has 2391 words, WoS Criteria is based on 32 words, and Open Alex's ML model is trained based on results from 151 Keywords further highlighting the differences in approaches of each database. These observations conform with the earlier seen presence of variation in SDG mapping of records across the databases by some previous studies (Armitage *et al.*, 2020, Purnell, 2022, Wang *et al.*, 2023). A more detailed analysis of the exact operation of each approach would reveal a further more detailed picture.

The mapping of records across the three databases shows differences in SDG classification linkages further reinforcing the finding as above. *While approaches used by Scopus and WoS have higher affinity of characterising/mapping records in multiple SDGs, in case of Open Alex, most records are assigned to a single SDG.* WoS has divergence values ranging from 1.1 to 3.4 with about 17% of records assigned to two or more SDGs. For Scopus this range is 1.5 to 3.2, with about 24% of records mapped to two or more SDGs. However, in Open Alex the divergence values range from 1.0 to 1.3 with only 1% records having been mapped to more than one SDGs. This suggests that Open Alex is more focused and conclusive to favour individual SDGs as compared to the other two databases.

Finally, *the network maps show the interlinkages between different SDGs by looking at the individual records and their mapping with multiple SDGs.* These linkages also vary across databases indicating differences in establishing linkages between the SDGs. This is possibly a result of the fact that individual SDG classifications are drawn separately without much consideration to the commonalities between them based on their targets and application areas.

Conclusion

This study has presented a detailed analysis of the variations that exist between the SDG mapping of records in three databases, namely, WoS, Scopus and Open Alex. Variations at the level of record volume, distribution across SDGs, divergence in mapping and linkages between SDGs are explored and presented. It is observed that the approach used by Open Alex classified publication records under fewer SDGs as compared to Scopus and WoS. This is indicated by the divergence score computed using the approach proposed in this article. There are also significant differences in SDG mapping linkages across the three databases. The analysis suggests that there are not only coverage level variations across the three databases, but there are also more methodological differences in SDG mapping schemes of the databases. The results of the present study bring out more detailed insight into the SDG mapping in the three databases.

The present study, however, has some limitations as well. It uses publication data for research output for the year 2023 from one country 'India'. However, owing to a significantly large scale of analysed data in the study, it is likely that the overall findings in terms of variability in SDG mapping across databases would still be visible in other large data samples. Further, only three databases are compared, the quantitative as well as the qualitative results are indicative of variations in these databases. It may be useful and interesting to study such variations in other databases (such as Dimensions, Google Scholar, Lens.org etc.) as well. Already, some studies have proposed alternative modes of classification of publication records into different SDGs (Wulff et al., 2023). Additionally, there are no studies to evaluate the accuracy of mapping publication records to SDGs at a more microscopic level, and therefore such a study can be conducted through a user-based annotation and evaluation.

References

Armitage, C. S., Lorenz, M., & Mikki, S. (2020). Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results? *Quantitative Science Studies*, **1(3)**, 1092-1108.

Bautista-Puig, N., Aleixo, A. M., Leal, S., Azeiteiro, U., & Costas, R. (2021). Unveiling the research landscape of sustainable development goals and their inclusion in higher education institutions and research centers: major trends in 2000–2017. *Frontiers in Sustainability*, *2*, 620743.

Hajikhani, A., & Suominen, A. (2022). Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. *Scientometrics*, *127*(11), 6661-6693.

IISD, Sustainable Development, International Institute for Sustainable Development (IISD), (2021). Accessed on May 27, 2023 from https://www.iisd.org/about-iisd/sustainable-development.

Kashnitsky, Y., Roberge, G., Mu, J., Kang, K., Wang, W., Vanderfeesten, M., ... & Labrosse, I. (2024). Evaluating approaches to identifying research supporting the United Nations Sustainable Development Goals. *Quantitative Science Studies*, 1-18.

Moyer, J. D., & Hedden, S. (2020). Are we on the right path to achieve the sustainable development goals?. *World Development*, *127*, 104749.

Ottaviani, M., & Stahlschmidt, S. (2024). On the performativity of SDG classifications in large bibliometric databases. *arXiv preprint arXiv:2405.03007*.

Purnell, P. J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case study of SDG 13—Climate Action. *Quantitative Science Studies*, **3(4)**, 976-1002.

Sachs, J. D. (2012). From millennium development goals to sustainable development goals. *The lancet*, **379**(9832), 2206-2211.

Singh, A., Kanaujia, A., Singh, V. K., & Vinuesa, R. (2024). Artificial intelligence for Sustainable Development Goals: Bibliometric patterns and concept evolution trajectories. Sustainable Development, 32(1), 724-754.

Singh, A., Kanaujia, A., & Singh, V. K. (2022). Research on sustainable development goals: how has indian scientific community responded?. *Journal of Scientific & Industrial Research*, *81*(11), 1147-1161.

Sweileh, W. M. (2020). Bibliometric analysis of global scientific literature on vaccine hesitancy in peer-reviewed journals (1990–2019). *BMC public health*, *20*, 1-15.

United Nations (2020). The Sustainable Development Goals Report 2020, Accessed from https://sdgs.un.org/publications/sustainable-development-goals-report-2020-24686 on 17 April 2025.

Wang, W., Kang, W., & Mu, J. (2023). Mapping research to the sustainable development goals (sdgs). *Preprint]. https://doi. org/10.21203/rs, 3*.

Wulff, D. U., Meier, D. S., & Mata, R. (2024). Using novel data and ensemble models to improve automated labeling of Sustainable Development Goals. *Sustainability Science*, 1-15.

Appendix

Records linked to each SDG in different databases

Table A1. Web of Science

WoS	SDG 01	SDG 02	SDG 03	SDG 04	SDG 05	SDG 06	SDG 07	SDG 08	SDG 09	SDG 10	SDG 11	SDG 12	SDG 13	SDG 14	SDG 15	SDG 16	SDG 17	Sum	TP_WoS	Divergence Score
SDG 01	568	251	267	0	158	0	0	103	134	223	0	0	73	0	63	5	14	1859	568	3.27
SDG 02	251	5557	1001	77	30	430	10	0	67	0	339	511	2890	885	1657	0	0	13705	5557	2.47
SDG 03	267	1001	45316	44	1320	1043	10	13	77	158	1253	151	719	762	364	65	0	52563	45316	1.16
SDG 04	0	77	44	564	2	0	0	0	3	11	0	80	80	77	77	0	0	1015	564	1.80
SDG 05	158	30	1320	2	1444	0	0	13	62	133	0	0	0	0	0	42	0	3204	1444	2.22
SDG 06	0	430	1043	0	0	5835	508	14	118	0	1026	940	1119	1872	481	4	0	13390	5835	2.29
SDG 07	0	10	10	0	0	508	10539	243	506	0	884	1084	1233	71	73	0	0	15161	10539	1.44
SDG 08	103	0	13	0	13	14	243	701	322	96	14	243	243	14	14	0	118	2151	701	3.07
SDG 09	134	67	77	3	62	118	506	322	2341	72	387	619	576	138	81	0	77	5580	2341	2.38
SDG 10	223	0	158	11	133	0	0	96	72	384	7	0	7	0	0	2	9	1102	385	2.86
SDG 11	0	339	1253	0	0	1026	884	14	387	7	6015	841	1574	980	430	4	0	13754	6015	2.29
SDG 12	0	511	151	80	0	940	1084	243	619	0	841	5313	1053	566	479	0	0	11880	5313	2.24
SDG 13	73	2890	719	80	0	1119	1233	243	576	7	1574	1053	9703	2334	2542	4	0	24150	9703	2.49
SDG 14	0	885	762	77	0	1872	71	14	138	0	980	566	2334	4020	1538	4	0	13261	4020	3.30
SDG 15	63	1657	364	77	0	481	73	14	81	0	430	479	2542	1538	4275	4	0	12078	4275	2.83
SDG 16	5	0	65	0	42	4	0	0	0	2	4	0	4	4	4	157	5	296	157	1.89
SDG 17	14	0	0	0	0	0	0	118	77	9	0	0	0	0	0	5	208	431	208	2.07

Table A2. Scopus

Scopus	SDG 01	SDG 02	SDG 03	SDG 04	SDG 05	SDG 06	SDG 07	SDG 08	SDG 09	SDG 10	SDG 11	SDG 12	SDG 13	SDG 14	SDG 15	SDG 16	SDG 17	Sum	TP_Scopus	Divergence Score
SDG 01	846	145	158	65	127	61	39	336	94	321	69	72	90	14	30	66	179	2712	846	3.21
SDG 02	145	5005	585	48	81	458	194	1232	317	105	173	953	828	131	420	47	404	11126	5005	2.22
SDG 03	158	585	31587	140	330	1232	473	382	674	367	871	533	383	331	377	216	328	38967	31587	1.23
SDG 04	65	48	140	1202	106	44	35	185	135	133	47	109	45	9	39	64	171	2577	1202	2.14
SDG 05	127	81	330	106	1145	22	8	206	61	275	37	21	24	7	7	266	105	2828	1145	2.47
SDG 06	61	458	1232	44	22	6459	769	408	1006	44	650	760	456	380	713	53	296	13811	6459	2.14
SDG 07	39	194	473	35	8	769	13881	537	1361	51	441	943	1663	99	228	26	501	21249	13881	1.53
SDG 08	336	1232	382	185	206	408	537	4101	880	349	347	1293	942	101	298	140	984	12721	4101	3.10
SDG 09	94	317	674	135	61	1006	1361	880	8326	138	590	1822	992	162	320	76	867	17821	8326	2.14
SDG 10	321	105	367	133	275	44	51	349	138	1572	87	85	87	18	29	128	366	4155	1572	2.64
SDG 11	69	173	871	47	37	650	441	347	590	87	4459	705	505	162	236	85	360	9824	4459	2.20
SDG 12	72	953	533	109	21	760	943	1293	1822	85	705	5172	935	198	405	84	816	14906	5172	2.88
SDG 13	90	828	383	45	24	456	1663	942	992	87	505	935	4915	194	341	49	643	13092	4915	2.66
SDG 14	14	131	331	9	7	380	99	101	162	18	162	198	194	1856	230	16	108	4016	1856	2.16
SDG 15	30	420	377	39	7	713	228	298	320	29	236	405	341	230	2933	47	255	6908	2933	2.36
SDG 16	66	47	216	64	266	53	26	140	76	128	85	84	49	16	47	1234	130	2727	1234	2.21
SDG 17	179	404	328	171	105	296	501	984	867	366	360	816	643	108	255	130	3033	9546	3033	3.15

Table A3. OpenAlex

Open Alex	SDG 01	SDG 02	SDG 03	SDG 04	SDG 05	SDG 06	SDG 07	SDG 08	SDG 09	SDG 10	SDG 11	SDG 12	SDG 13	SDG 14	SDG 15	SDG 16	SDG 17	Sum	TP_OpenAlex	Divergence Score
SDG 01	3868	24	77	0	2	1	0	23	0	8	2	0	0	0	0	0	0	4005	3867	1.04
SDG 02	24	21832	44	3	24	17	0	0	0	0	1	91	46	1	64	0	0	22147	21830	1.01
SDG 03	77	44	49993	16	63	0	18	0	0	0	4	0	0	13	2	2	17	50249	49951	1.01
SDG 04	0	3	16	9841	31	0	0	2	1	19	2	0	0	1	0	10	0	9926	9837	1.01
SDG 05	2	24	63	31	5763	0	0	47	0	14	0	0	0	1	0	39	0	5984	5763	1.04
SDG 06	1	17	0	0	0	14894	67	0	1	0	2	3	2	126	27	0	0	15140	14887	1.02
SDG 07	0	0	18	0	0	67	33864	0	7	0	6	3	22	0	0	0	0	33987	33855	1.00
SDG 08	23	0	0	2	47	0	0	6458	10	15	1	6	0	3	0	3	1	6569	6457	1.02
SDG 09	0	0	0	1	0	1	7	10	13049	0	16	122	0	0	0	0	69	13275	13049	1.02
SDG 10	8	0	0	19	14	0	0	15	0	5882	1	0	0	1	0	382	6	6328	5882	1.08
SDG 11	2	1	4	2	0	2	6	1	16	1	7849	0	45	18	30	0	0	7977	7848	1.02
SDG 12	0	91	0	0	0	3	3	6	122	0	0	8101	0	2	539	0	0	8867	8101	1.09
SDG 13	0	46	0	0	0	2	22	0	0	0	45	0	5058	23	6	0	0	5202	5058	1.03
SDG 14	0	1	13	1	1	126	0	3	0	1	18	2	23	7353	33	1	0	7576	7351	1.03
SDG 15	0	64	2	0	0	27	0	0	0	0	30	539	6	33	10496	1	0	11198	10494	1.07
SDG 16	0	0	2	10	39	0	0	3	0	382	0	0	0	1	1	8850	0	9288	8850	1.05
SDG 17	0	0	17	0	0	0	0	1	69	6	0	0	0	0	0	0	4028	4121	4028	1.02