Estimating transformative agreement impact on hybrid open access: A comparative large-scale study using Scopus, Web of Science and open metadata

Najko Jahn[1][*] (0000-0001-5105-1463)

[1] Göttingen State and University Library, University of Göttingen, Germany.

[*] Correspondence: najko.jahn@sub.uni-goettingen.de

Abstract

This study compares open metadata from hoaddata, an openly available dataset based on Crossref, OpenAlex and the cOAlition S Journal Checker Tool, with proprietary bibliometric databases Scopus and Web of Science to estimate the impact of transformative agreements on hybrid open access publishing. Analysing over 13,000 hybrid journals between 2019-2023, the research found substantial growth in open access due to these agreements, although most articles remain paywalled. The results were consistent across all three data sources, showing strong correlations in country-level metrics despite differences in journal coverage and metadata availability. By 2023, transformative agreements enabled the majority of hybrid open access articles, with particularly high adoption in European countries. The analysis revealed strong alignment between first and corresponding authorship when measuring agreement uptake by publisher and country. This comparative approach supports the use of open metadata for large-scale hybrid open access studies, while using multiple data sources together provides a more robust understanding of hybrid open access adoption than any single database can offer, overcoming individual limitations in coverage and metadata quality.

Keywords: bibliometric data sources, open metadata, hybrid open access, transformative agreements, curative bibliometrics

1   Introduction

Transformative agreements are a much-discussed library licensing model for transitioning subscription-based journal publishing to full open access. Although these agreements can vary considerably, they are mainly aimed at hybrid journal bundles, allowing authors from participating institutions to publish open access in these journals while providing reading access to the entire portfolio. Library consortia and funders have begun to coordinate how these agreements are negotiated, managed and evaluated. At the same time, efforts are underway to improve transparency through open data on contracts and payments. However, this information is still only partially available, which limits large-scale studies of the impact of transformative agreements on the transition of hybrid journals to full open access. By comparing open metadata from hoaddata, an openly available dataset on hybrid open access based on metadata from Crossref, OpenAlex and the cOAlition S Journal Checker Tool, with the established bibliometric databases Scopus and Web of Science, this study aims to provide a robust estimate of the impact of transformative agreements on hybrid open access. This comparative approach informs bibliometric research and practice by highlighting the strengths and weaknesses of each data source individually and collectively.

Since their initial proposal by the Max Planck Digital Library (Schimmer et al., 2015), transformative agreements have led to rapid growth in the number of negotiated agreements and open access articles (Jahn, 2025; McCabe & Mueller-Langer, 2024; Rothfritz et al., 2024). Before, only a small proportion of articles in hybrid journals were openly available (Björk, 2012; Laakso & Björk, 2016; Piwowar et al., 2018), with publications from only a few European countries with dedicated funding policies and the high-energy physics SCOAPˆ3 consortium standing out (Huang et al., 2020; Kohls & Mele, 2018; Pinfield, 2015; Robinson-Garcia et al., 2020). Related coordination activities, in particular through the OA2020 and ESAC (Efficiency and Standards for Article Charges) initiatives, have also led to improved workflows between library consortia and publishers (Campbell et al., 2022; Geschuhn & Stone, 2017). However, critics have raised concerns about these agreements, particularly regarding market concentration (Butler et al., 2023; Shu & Larivière, 2023), the failure of most hybrid journals to convert to full open access (Kiley, 2024; Momeni et al., 2021), reduced competition and incentives (McCabe & Mueller-Langer, 2024; Schmal, 2024), and the continuation of publication fees that widen gaps between well-resourced and under-resourced institutions (Babini et al., 2022; Ross-Hellauer et al., 2022)

While consortia evaluations confirm the growth of open access through transformative agreements and acknowledge increasing coordination, they reach mixed policy conclusions. For instance, the British Jisc evaluation (Brayman et al., 2024) found that the consortium's transformative agreements had a significant impact on national open access growth but a limited effect on the global open access transition, suggesting a need to reevaluate their strategy. Similar policy reconsiderations

emerged in Norway and Sweden (Holden et al., 2023; Widding, 2024). Meanwhile, the funder initiative cOAlition S recommended ending financial support for these agreements in 2024, but still considers hybrid open access to be compliant with funders' open access policies. Conversely, the German DEAL consortium extended agreements with major publishers until 2028, while also outside of Europe, more transformative agreements are being put in place (for instance in Latin America, as described in Muñoz-Vélez et al. (2024)). Responding to criticism, the Max Planck Digital Library recently outlined strategies to evolve the original model, focusing on achieving fully open access while controlling costs and addressing identified weaknesses (Dér, 2025)

Despite controversies, transformative agreements can benefit the landscape of open research information as library consortia negotiate for comprehensive open metadata, helping avoid data analytics from becoming a commodity controlled by publishers (Aspesi & Brand, 2020; McCabe & Mueller-Langer, 2024). For instance, ESAC recommends using Crossref (Hendricks et al., 2020) to share metadata supporting workflows including open access licenses and funding information (Geschuhn & Stone, 2017). Simultaneously, the bibliometric community pushes for more open metadata through this publisher-driven DOI registration platform and continuously monitors the progress (N. J. van Eck & Waltman, 2022).

Many open scholarly data services that build on Crossref have benefited from these developments. A prominent example is OpenAlex, which has gained attention for its inclusivity, openness and comprehensive coverage of publication metadata (Priem et al., 2022). In a recent evaluation, Céspedes et al. (2025) reviewed the current evidence base on the strengths and weaknesses of OpenAlex. They highlight both its increasing use in bibliometric research and its adoption by research organisations. Comparative studies have generally found OpenAlex to be a reliable alternative to commercial databases (Alperin et al., 2024; Culbert et al., 2025), although data quality issues occurred for key metadata fields required for open access analysis, including evidence of open access status itself (Jahn et al., 2023; Simard et al., 2024), as well as authorships (Culbert et al., 2025), affiliations (Zhang et al., 2024), author roles (Jahn, 2025) and document types used to distinguish between different article types (Haupka et al., 2024). However, these studies highlight the rapid improvements and new possibilities of combining open metadata and the curation efforts of OpenAlex in conjunction with the growing awareness of the importance of open metadata, as demonstrated by the Barcelona Declaration on Open Research Information.

Consequently, there is growing interest in using open bibliometric databases such as Crossref and OpenAlex to assess the impact of transformative agreements. Similar to established proprietary bibliometric databases such as Scopus and Web of Science, these data sources lack specific data points, particularly about open access payments, which are typically not shared (Haustein et al., 2024). However, bibliometric data can be enriched with open metadata about journals and institutions

involved in transformative agreements (Jahn, 2025). Jonge et al. (2025) demonstrated this potential by combining data from the cOAlition S Journal Checker tool and OpenAlex, achieving 89% accuracy when validated against Dutch research information. While open data approaches show promise, understanding how these results compare when the same methodology is applied to proprietary databases such as Web of Science and Scopus remains unexplored.

This study compares open and proprietary data sources for evaluating transformative agreements. It adapts the approach from Jahn (2025), which combines data from the cOAlition S Journal Checker tool, Crossref and OpenAlex, and applies the same methods to Web of Science and Scopus data. This comparative approach allows the evaluation of different bibliometric data sources to monitor transformative agreements, while assessing their respective strengths and limitations. Specifically, this study compares the indexing of more than 13,000 hybrid journals across these sources, and examines whether consistent results can be obtained in terms of open access adoption between 2019 and 2023, analysed by publishers and countries. By comparing results across these data sources, this large-scale study provides a more robust understanding of the impact of transformative agreements on the transition to full open access. It also reveals the comparative strengths and weaknesses of different bibliometric databases in measuring scholarly publishing's shift toward openness.

2   Data and methods

As shown in Figure 1, the methodology involved comparing hoaddata, an openly available collection of open research information on hybrid open access, with the bibliometric databases Web of Science and Scopus. This section introduces the initial data sources, followed by a presentation of the necessary data processing steps to obtain eligible articles enabled by transformative agreements using author roles (first and corresponding) and harmonised affiliation data.
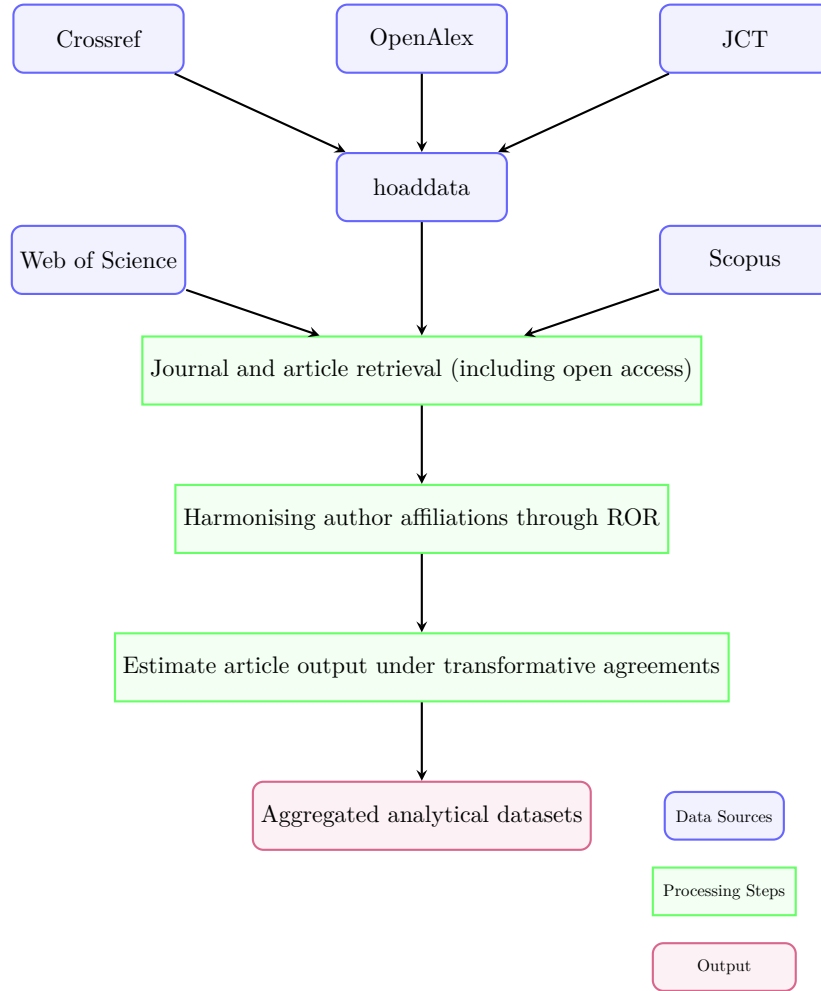
2.1   Data sources

hoaddata.   hoaddata, developed and maintained by the author to support open access monitoring and research (e.g. Jahn (2025)), is an R data package that regularly collects information on hybrid open access uptake from multiple openly available data sources (Jahn, 2024). It combines article-level metadata from Crossref (Hendricks et al., 2020) and affilation metadata from OpenAlex (Priem et al., 2022) with transformative agreement information from the Transformative Agreement Data dump used by the cOAlition S Journal Checker Tool (JCT)[1], which links journal and institutional data about participating research organisations to agreements in the ESAC registry.

hoaddata followed good practices for computational reproducibility using R (Marwick et al., 2018). The package, which includes data, code, a test suite and documentation, is openly available on GitHub. To ensure computational reproducibility while aggregating the data, a GitHub Actions

---

[1] https://journalcheckertool.org/transformative-agreements/

Figure 1. Data processing workflow for comparing hybrid open access uptake across bibliometric data sources. The workflow shows how data from different sources (hoaddata, derived from Crossref, OpenAlex, Transformative Agreement Data dump used by the cOAlition S Journal Checker Tool (JCT), the Web of Science, and Scopus) were processed to enable comparative analysis.

continuous integration and delivery (CI/CD) workflow handles data retrieval from the SUB Göttingen's open scholarly data warehouse based on Google BigQuery, which provides high-performant programmatic access to monthly snapshots of Crossref and OpenAlex.[2] The workflow has run regularly to fetch updates from these data sources since 2022. The package version used in this study is 0.3, containing data from the Crossref 2024-08 dump provided to Crossref Metadata Plus subscribers and the OpenAlex 2024-08-29 monthly dump. It covers agreements collected between July 2021 to July 2024 from the JCT. This version including the computation log is available on GitHub (https://github.com/subugoe/hoaddata/releases/tag/v.0.3).

Web of Science. Clarivate Analytics' Web of Science (WoS) is a well-established proprietary bibliometric database consisting of several collections (Birkle et al., 2020). Web of Science

---

[2] https://subugoe.github.io/scholcomm_analytics/data.html

is a selective database with a focus on base research (Stahlschmidt & Stephen, 2022; Visser et al., 2021). The collections considered in this study were the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (AHCI).

These collections provide important data points for analysing open access: author affiliations and roles, differentiation of journal articles into document types representing different types of journal contributions, such as original articles or reviews, and open access status information derived from OurResearch's Unpaywall (Piwowar et al., 2018), the same provider as OpenAlex. However, Web of Science lacks information about journals and articles under transformative agreements.

For programmatic access to article-level data, this study used the database of the Kompetenznetzwerk Bibliometrie (KB) in Germany. The KB processes raw XML data provided by Clarivate Analytics, which are ingested into an in-house PostgreSQL database under a uniform schema. To support reproducibility, KB maintains annual snapshots of the database. Accordingly, this study used the annual snapshot from April 2024 (wos_b_202404), which is considered to cover almost the entire previous publication year (Schmidt et al., 2024).

Scopus. Elsevier's Scopus, launched in 2004, is another widely used proprietary bibliometric database for measuring research (Baas et al., 2020). Similar to Web of Science, Scopus is selective with regard to the journals it indexes. However, its journal coverage is much broader than that of the Web of Science collections considered in this study, as it indexes a wider range of applied research journals (Singh et al., 2021; Stahlschmidt & Stephen, 2022; Visser et al., 2021). With detailed metadata about article types, open access status information derived from Unpaywall, author roles, and disambiguated affiliations, Scopus also contains important data to assess open access uptake, although no direct information regarding transformative agreements was available at the time of the study.

This study used the Scopus annual snapshot from April 2024 as provided by KB (scp_b_202404). The same KB curation effort as that for Web of Science was applied to Scopus (Schmidt et al., 2024).

2.2   Data processing steps

Determining the hybrid journal publication volume. Following Jahn (2025), the starting point was a unified dataset of several safeguarded JCT snapshots[3]. JCT journal data were enriched with ISSN variants linked to ISSN-L. To identify hybrid journals, a comprehensive exclusion of fully open access journals was performed using multiple journal lists including the Directory of Open Access Journals (DOAJ). The resulting hybrid journal data were made available via hoaddata and were used to independently determine the publication volume for each database.

hoaddata relies on Crossref to obtain journal publication volume and open access status through Creative Commons (CC) licence information relative to the published version ("version of

---

[3] https://github.com/njahn82/jct_data

record"). The article metadata included DOIs, publication dates, open access information, and author roles and affiliations. Publication years were determined using the earliest known date of publication in a journal. In hoaddata, this corresponded to Crossref's issued date. For Web of Science and Scopus, the earliest publication date was used where available, with Scopus dates specifically determined by the KB through version tracking of the raw data.

Many transformative agreements typically cover only certain types of journal articles, in particular original research articles including reviews (Borrego et al., 2021). Because of limited information on these document types in open scholarly data (Haupka et al., 2024), hoaddata used an extended version of Unpaywall's paratext recognition approach to exclude non-scholarly content (Piwowar et al., 2018). To exclude conference supplements, which are often not covered by transformative agreements, only articles published in regular issues, indicated by numerical pagination, were considered. For Web of Science and Scopus, their established and mainly accurate document type classifications (Donner, 2017; Maisano et al., 2025) were used to identify original research articles and reviews, referred to as original articles throughout this study.

Identifying open access articles in hybrid journals.   Articles in hybrid journals were considered open access when publishers made them freely available under a CC license on their platforms. While hoaddata obtained this information from Crossref license metadata, Web of Science and Scopus relied on Unpaywall as evidence source. Unpaywall also uses Crossref license metadata, but supplements them by parsing publisher websites directly, addressing cases where publishers do not provide machine-readable CC license information (Piwowar et al., 2018). This additional parsing remains necessary despite transformative agreement workflows recommending the deposition of CC licenses during DOI registration (Geschuhn & Stone, 2017). Both Web of Science[4] and Scopus[5] defined hybrid open access consistently as content available under CC licenses on publisher platforms, distinguishing it from bronze open access that lack such explicit license information, or use publisher-specific licenses (Piwowar et al., 2018).

Harmonising author affiliations across databases.   Author affiliations were retrieved for both the first and, if available, the corresponding authors to prepare the linking between articles and institutions covered by transformative agreements. To improve data retrieval, JCT institution data were enriched with ROR-IDs from associated institutions, such as university hospitals or institutes of large research organisations such as the Max Planck Society, according to OpenAlex's institution entity. Database-specific affiliation identifiers were used to handle different address variants: ROR-IDs from OpenAlex for hoaddata, affiliation enhanced names for Web of Science, and Scopus Affiliation Identifier. Additionally, ISO country codes were retrieved for each author's address to compile

---

[4] `https://webofscience.help.clarivate.com/en-us/Content/open-access.html`

[5] `https://blog.scopus.com/posts/scopus-filters-for-open-access-type-and-green-oa-full-text-access-option`

country-level statistics.

Because neither Web of Science nor Scopus supported ROR-ID at the time of data retrieval, the institution identifier used by the JCT, a two-step matching process was implemented to harmonise the affiliation data. First, 2,782,540 articles from 6,457 institutions with ROR-IDs in the JCT data since 2017 (according to hoaddata) were processed to map the first authors' ROR-IDs to corresponding proprietary affiliation identifiers in Web of Science and Scopus using DOI matching. Then, an algorithm selected the most frequent ROR ID and proprietary identifier pairs to handle multiple affiliations and organisational hierarchy differences.

This process linked 6,375 ROR-IDs to 4,894 Scopus Affiliation IDs and 6,034 ROR-IDs to 2,422 enhanced affiliation strings in Web of Science. Quality evaluation through random sampling of 50 pairs revealed an error rate of 22% for Web of Science (11 mismatches) and 6% for Scopus (three mismatches). Upon inspection, these mismatches primarily occurred with less-represented institutions having only a few publications, introduced through multiple affiliations of single authors. The difference between databases suggests that Scopus' affiliation control aligns more closely with ROR than with Web of Science.

Estimating open access in hybrid journals covered by transformative agreements. Based on these compiled matching tables, articles eligible under transformative agreements could also be obtained from Web of Science and Scopus, although they did not contain ROR-IDs used by the JCT. The estimation of eligible articles followed Jahn (2025) and included matching of both journals and participating institutions according to the JCT. The matching also considered the duration of agreements according to the ESAC registry, with only those matches where an agreement was actually in place being considered for subsequent analysis. A related study (Jonge et al., 2025), applied to publications funded by the Dutch Research Council (NWO) and validated against internal invoicing data, confirmed that such matching can accurately identify most articles under transformative agreements.

2.3   Data records

As a result of the comprehensive data processing described above, datasets on open access in hybrid journals included in transformative agreements were aggregated for each database at the country and journal levels by year. Table 1 provides a general overview of coverage between 2019 and 2023. It shows that the majority of hybrid journals published at least one original research article or review, collectively marked as "original article" in the following, during the five-year period. These journals have provided the basis for the subsequent calculation of article-level indicators.

While hoaddata only covered articles with a DOI, Scopus and Web of Science publication indicators were calculated using their database identifier. A subsequent comparison of DOI coverage shows that non-original articles in Web of Science often lacked a DOI. This was particularly the case

Table 1

Coverage of hybrid journals in transformative agreements 2019-23.

	hoaddata*	Web of Science	Scopus
Hybrid journal coverage			
Active journals	12,890	8,655	11,888
Active journals (core)	12,888	8,655	11,878
Active journals (core) with OA	11,348	8,392	11,313
Publication volume			
Total published articles	9,740,015	8,616,053	8,117,644
Original articles	8,158,425	6,708,083	7,317,703
Digital Object Identifier (DOI) coverage			
Articles with DOI	9,740,015	7,713,796	8,105,112
Original articles with DOI	8,158,425	6,695,661	7,314,327
Open Access (OA) metrics			
OA articles	998,699	1,112,758	974,099
Original OA articles	969,817	1,019,784	922,578
Original articles with affiliation data			
First author articles	7,242,542	6,294,855	7,232,017
Corresponding author articles	5,534,207	6,291,441	6,898,487

* Journal article metadata from Crossref, except affiliations from OpenAlex

for meeting abstracts, which are notably prevalent in Health Sciences journals (Melero-Fuentes et al., 2025) and are not indexed by Scopus (Donner, 2017). Open access indicators were aggregated by DOI, because Unpaywall only collects information on open access status for articles with DOI. A closer look at original articles with affiliation data, in line with related research (Zhang et al., 2024), reveals a lack of affiliation data in the case of OpenAlex, the affiliation data source used by hoaddata, compared to Web of Science and Scopus. In particular, only approximately two-thirds of the articles examined provided corresponding author affiliations. The proportion for first authors was 89%. At the time of writing, OpenAlex disclosed limited coverage of corresponding authorship data [6]. Therefore, only first author data for hoaddata were considered in the following analysis.

---

[6] `https://docs.openalex.org/api-entities/works/work-object/authorship-object#is_corresponding`

2.4 Data analysis

TTidyverse tools (Wickham et al., 2019) for the R programming language (R Core Team, 2024) were used throughout this largely automated data collection and analysis process. Rank correlations were calculated using the Hmisc package (Harrell Jr, 2003). Source code analysis, including all queries used to obtain the data, is available on GitHub (`https://github.com/njahn82/hoa_validation`). An interactive supplement exploring correlations between the data sources examined by country and publisher is available via HuggingFace Spaces: `https://huggingface.co/spaces/najkoja/hoa_replication`.

3 Results

This section first presents the indexing of hybrid open access by comparing open data sources with the proprietary bibliometric databases Scopus and Web of Science. Then, using the same methods, indicators at the publisher and country levels are calculated independently for each database and compared to assess the suitability of the bibliometric databases for investigating transformative agreements.

3.1 Coverage comparision

Overview. Figure 2 presents the coverage of hybrid journals included in transformative agreements, based on the intersections of the examined data sources. The intersection sets of journals and articles are visualised as an UpSet graph (Krassowski, 2020; Lex et al., 2014). The analysis included hybrid journals that published at least one open access article between 2019 and 2023, based on open access status information from each database. Only original articles were included in the analysis.

Journal coverage analysis revealed that 66% (n = 7,970) of the hybrid journals included in transformative agreements were indexed in all three databases (Figure 2A). The second largest set consisted of journals indexed in both hoaddata and Scopus, comprising 21% (n = 2,595) of hybrid journals. Notably, 6% (n = 739) of journals were exclusively contained in hoaddata, while another 6% (n = 748) were found only in Scopus. Of these, 354 were available in Web of Science. Upon inspection, this group of hybrid journals exclusively covered in proprietary data sources mainly represented hybrid journals for which no open access evidence could be retrieved from Crossref, the open access evidence source for hoaddata.

In terms of article coverage, Figure 2B shows the total publication volume per combination in terms of DOI availability. The largest set of hybrid journals covering all three data sources also represented the largest number of articles. In total, these journals recorded 6,289,687 articles, represented by the blue bars. They recorded 94% of the original articles with a DOI indexed in the Web of Science sample, and 86% in Scopus. Another 657,697 articles were exclusive to the intersection between Scopus and hoaddata. Exclusively in hoaddata were 177,110 articles, and exclusively in the
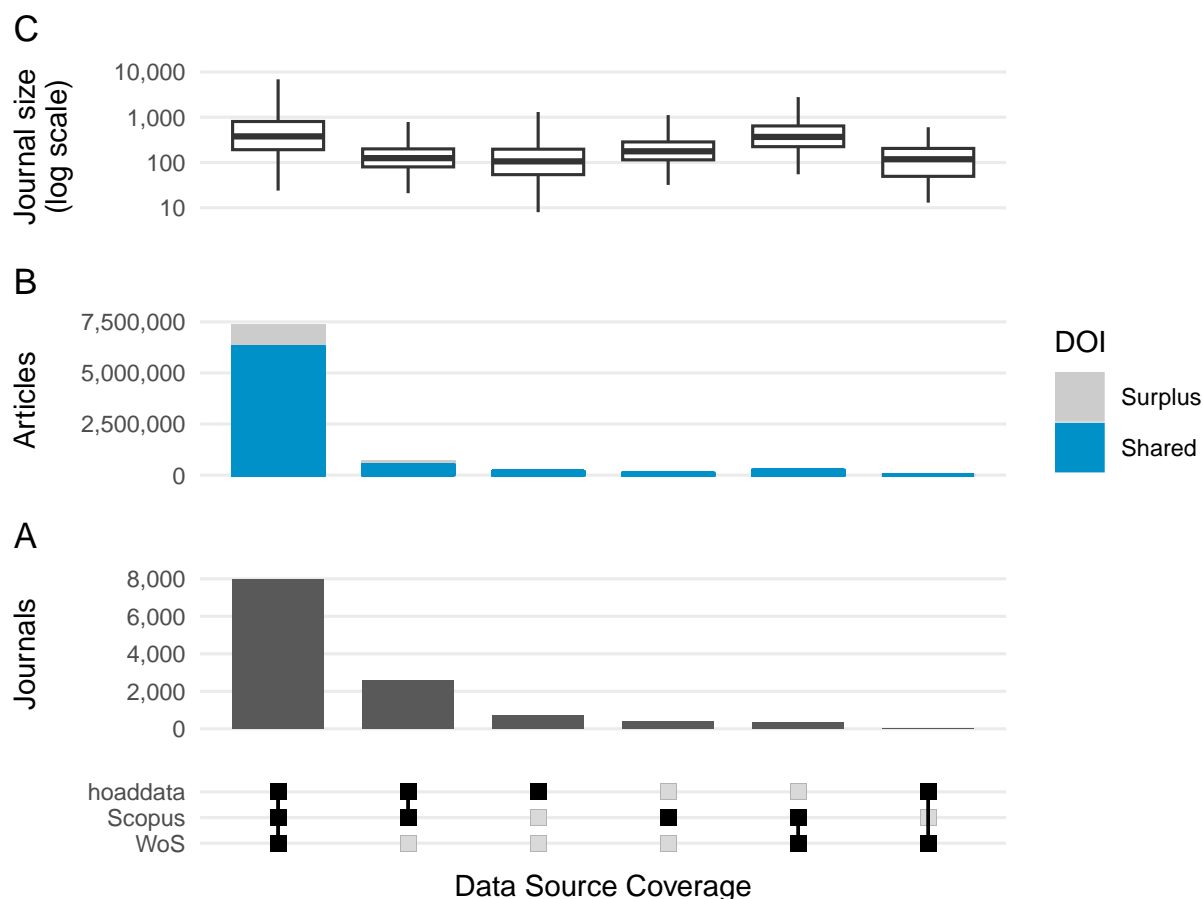
Figure 2. Comparison of hybrid journal indexing by data source, 2019-2023. Only hybrid journals present in the cOAlition S Transformative Agreement Data dump with at least one open access article are considered. A) presents the number of journals, B) the number of articles (DOI), distinguishing between shared DOI corpus and surplus in hoaddata. Box plots (C) shows the five-years journal article volume (log-scale).

proprietary databases were 325,194 articles.

Figure 2B also shows the surplus of articles with DOI that were only available via hoaddata (grey area). In the case of hybrid journals covered by all three data sources, 1,023,882 DOIs were present only in hoaddata. After validation at the DOI level using KB databases and manual inspection, the main reason for missing DOI coverage in the proprietary database was insufficient classification of journal content as original articles during the compilation of hoaddata. Particularly, letters and editorials could not be detected fully. Moreover, paratext recognition failed for 37% of DOIs to identify non-scholarly content, such as front matters or reviewer lists, which are generally not indexed by Scopus and Web of Science. To a lesser extent, differences in publication and indexing dates were reasons for non-overlapping DOIs.

Using overlapping DOIs, the publication volume between 2019 and 2023 was calculated for each journal. Figure 2C illustrates the distribution for each combination. It shows a large spread of

journal size across journals mutually covered by all three data sources. These journals published more on average than journals covered by only one or two data sources, with the exception of journals exclusively covered by Scopus and Web of Science. In particular, journals covered exclusively by hoaddata were substantially smaller. Upon inspection, these were often newly launched hybrid journals, which explains the relatively low five-year publication volume. An example is *Digital society* that published 86 articles. This hybrid journal was launched in 2022, and had since been covered by various Springer Nature transformative agreements.

Coverage by publisher portfolio. Figure 3 presents the coverage of hybrid journals in transformative agreements across data sources from 2019 to 2023 focusing on publisher portfolios. The analysis highlights the dominance of the three largest publishers, Elsevier, Springer Nature, and Wiley, which collectively accounted for 46% of the hybrid journals and 62% of the articles published during this five-year period. In terms of article volume, Elsevier led with 2,435,478 articles (34% of the total) published across 1,941 hybrid journals (16% of the total). Springer Nature followed with 1,237,700 articles (17%) in 2,294 hybrid journals, although it recorded the largest number of hybrid journals (19%). Wiley accounted for 814,349 articles (11%) in 1,337 hybrid journals (11%). The remaining 54 publishers collectively accounted for 2,733,994 articles (38%) in 6,442 hybrid journals (54%).

The three largest publishers, Elsevier, Springer Nature, and Wiley, were best represented at the intersection of all three data sources (hoaddata, Scopus, and Web of Science). Together, they comprised 4,347 hybrid journals (55% of the intersectional set) and dominated article coverage (n = 4,143,663; 66%) as determined through shared DOIs. When examining the publication volume per journal (Figure 3C), Elsevier published, on average, the largest journals, followed by Springer Nature and Wiley.

A comparison of publisher portfolios across different indexing sets demonstrates that publishers were not represented uniformly. Notably, Springer Nature exhibited 517 hybrid journals exclusively indexed in both hoaddata and Scopus. This set included journals from the Chinese Academy of Science, German-language medical journals, and Eastern European publications including the *Journal of Mathematical Sciences*, which also publishes English-language translations of Russian-language works. Additionally, this subset included titles with a broader disciplinary focus, such as *SN Computer Science*, and newly launched hybrid journals such as *Nature Computational Science*, which started in 2021 and were indexed in Scopus but not yet in Web of Science. The set also captures discontinued journals, providing further insights into the dynamics of journal publishing.

Examining publisher portfolios not covered by hoaddata, but present in Scopus or Web of Science, identified several publishers with missing Creative Commons (CC) license information in Crossref. In particular, Emerald represented 321 journals with 86,269 articles, AIP Publishing accounted for 24 journals with 64,898 articles, and World Scientific recorded 87 journals with 42,531 articles. In total, 9 publishers did not share CC licenses with Crossref and were therefore not
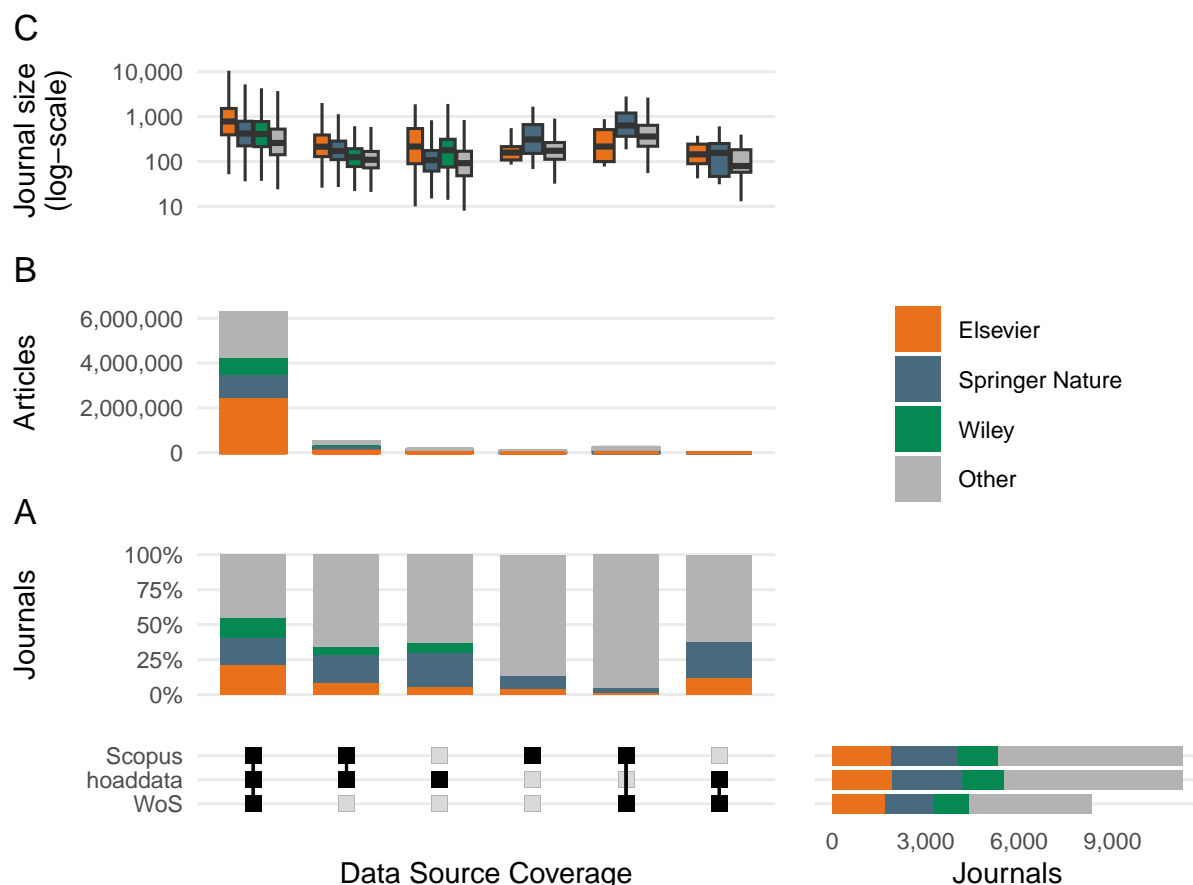
Figure 3. Comparison of hybrid journal indexing by data source and publisher, 2019-2023. Only hybrid journals present in the cOAlition S Transformative Agreement Data dump with at least one open access article are considered. A) presents the percentage of journals by publisher, B) the number of articles by publisher (shared DOI). Box plots (C) shows the five-years journal article volume (log-scale) by publisher.

represented in hoaddata.

An inspection of individual journals also uncovered discrepancies in Unpaywall's open access identification for certain publishers that typically share CC license metadata with Crossref. Notably, some subscription-only journals contained one or two articles erroneously tagged as hybrid open access by Unpaywall, which was subsequently reflected in Scopus and Web of Science. Examples of such misclassifications include Elsevier's *Journal of Bioscience and Bioengineering*, and Springer Nature's *Journal of Mechanical Science and Technology.*

3.2 Open Access Indicator Comparison

This section examines the uptake of open access in hybrid journals, focusing on the influence of transformative agreements across hoaddata, Scopus, and Web of Science. The aim was to assess whether consistent results could be derived from these data sources despite differences in coverage and methodologies. Following Jahn (2025), indicators were calculated for each data source and comprised

the number and proportion of open access articles, including those enabled by transformative agreements, from 2019 to 2023. For Web of Science and Scopus, the impact of transformative agreements was estimated using both first and corresponding authorships, while hoaddata indicator calculations were limited to first author affiliations.
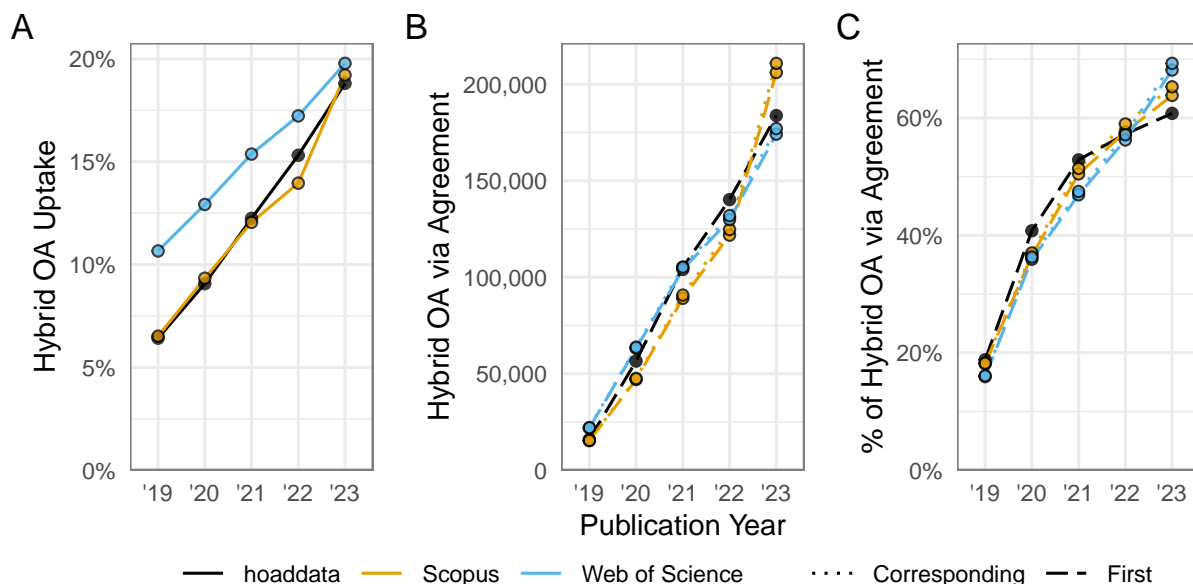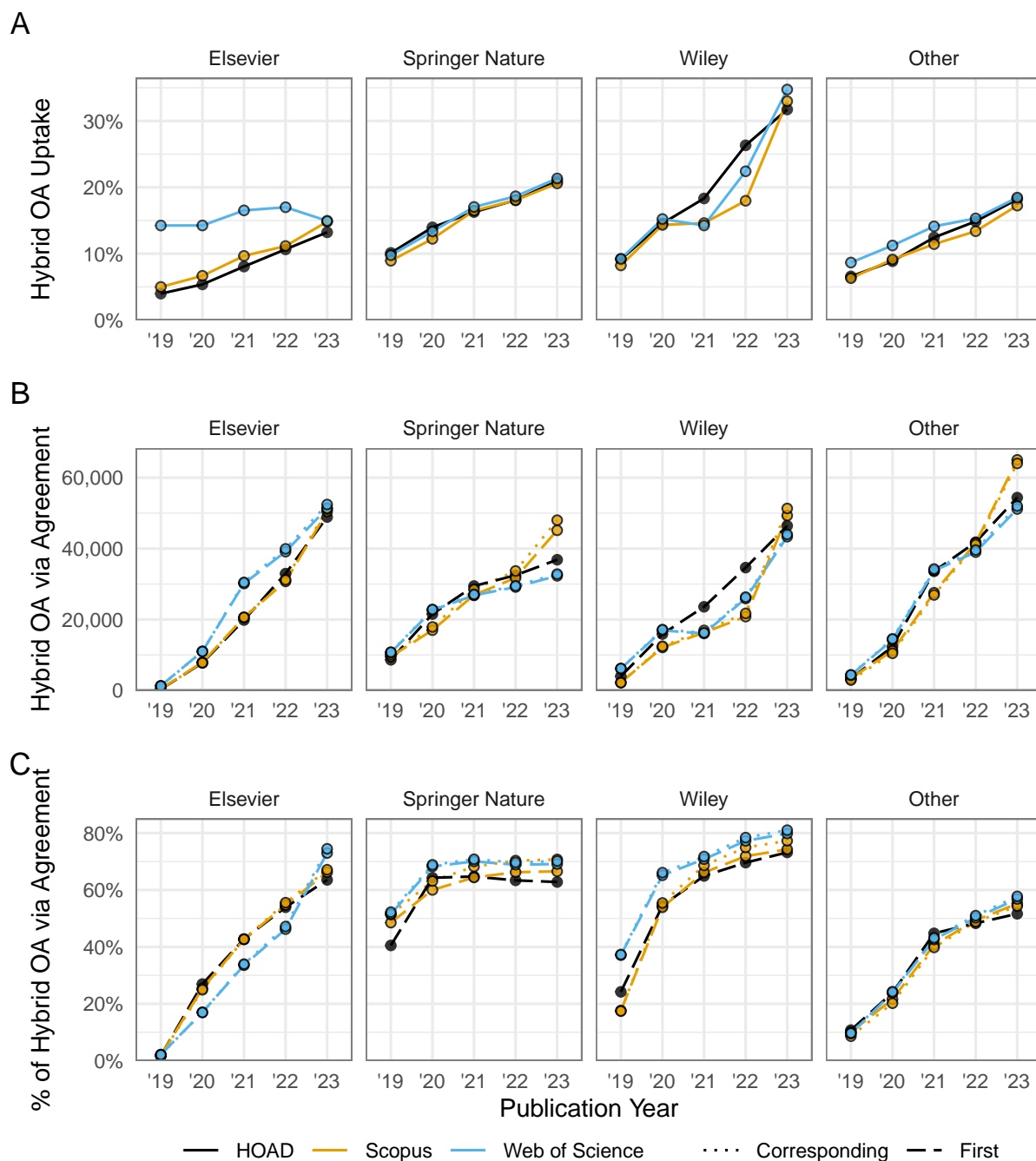

Figure 4. Development of open access in hybrid journals in transformative agreements by data source and author role, 2019-2023. Figure shows the open access percentage (A), the number (B) and the percentage (C) of open access articles enabled by transformative agreements.

Overview. Figure 4A shows a moderate growth of open access in hybrid journals, which is consistent across hoaddata (black line), Scopus (yellow line), and Web of Science (blue line). According to hoaddata, hybrid open access uptake increased from 6.4% (n = 85,071) in 2019 to 19% (n = 302,358) in 2023. Similarly, Scopus recorded a growth from 6.5% (n = 84,648) in 2019 to 19% (n = 322,850) in 2023. However, Web of Science recorded higher open access uptake in the early years, before converging to a similar level in 2023, from 11% (n = 137,202) in 2019 to 20% (n = 255,481) in 2023. This suggests a different approach towards labelling hybrid open access by Web of Science.

Similarly, hybrid open access by transformative agreements substantially increased between 2019 and 2023 (Figures 4B and C). Trends were consistent when measuring the first (dashed line) and corresponding authorship (dotted line). According to Scopus, 479,297 open access articles could be attributed to transformative agreements based on first author metadata (increasing from 15,341 to 206,084) and 489,262 using corresponding author metadata (from 15,444 to 210,816). Web of Science recorded 493,028 open access articles via transformative agreements using first author metadata (increasing from 21,871 to 174,126) and 500,076 using corresponding author metadata (from 22,092 to 177,030). hoaddata, lacking corresponding author data, linked 501,649 articles to transformative agreements, increasing from 16,010 in 2019 to 183,757 in 2023.

Since 2021 (hoadata, Scopus) resp. 2022 (Web of Science), transformative agreements have enabled the majority of open access articles in hybrid journals. For the first authors, the share ranged between 61% (hoaddata), 64% (Scopus), and 68 % (Web of Science) by 2023. For the corresponding authors, the shares were slightly larger, with Scopus recording 65% and Web of Science 69% by 2023. However, substantial hybrid open access was still facilitated outside transformative agreements, likely through APCs paid from discretionary research funds (Suber, 2012).

Figure 5. Development of open access in hybrid journals included in transformative agreements by data source, author role and publisher, 2019-2023. Figure shows the open access percentage (A), the number (B) and the percentage (C) of open access articles enabled by transformative agreements.

Open access by publishers. When considering open access trends by publisher (see Figure 5), the observed differences in early uptake rates between hoaddata and Scopus compared to Web of Science can be largely attributed to articles published in Elsevier hybrid journals, the largest publisher in our sample. Both hoadata and Scopus reported a steady increase in open access uptake between 2019 and 2023 (hoaddata from 4% to 13%; Scopus from 5% to 15%). In contrast, Elsevier's share remained relatively constant, increasing only slightly from 14% to 15% according to Web of Science. Upon inspection, this discrepancy was primarily due to articles under the publisher's open archive. These articles, made freely available after an embargo period under Elsevier's user license, were tagged as hybrid open access in Web of Science, even though its documentation[7] specified that only articles under a CC license variant were considered. Previous research (Haustein et al., 2024; Jahn et al., 2022) has shown that Elsevier provided a substantial portion of its articles under this license, explaining the relatively large and stable share of open access over the years.

Differences in open access evidence are also evident for Wiley. Specifically, Web of Science and Scopus recorded a drop in 2021 and 2022 compared to hoaddata. For these two years, hoaddata reported 32,445 more open access articles than Scopus and 29,854 more open access articles than Web of Science. This discrepancy is presumably due to challenges in fetching full-texts by Unpaywall, the open access evidence source for Scopus and Web of Science. According to Unpaywall's software version history, HTTP redirects from Wiley's publisher platform, preventing Unpaywall from parsing license information from full-texts.[8] hoaddata, which relies solely on Crossref metadata for open access identification, was unaffected by these issues.

Despite these differences in open access evidence, the three data sources show consistent temporal trends in hybrid open access enabled by transformative agreements (see Figure 5B and C). Wiley emerged as the fastest-growing publisher in terms of open access uptake, with more than 30% of articles in hybrid journals reported as open access in 2023 across the examined data sources, followed by Springer Nature. Elsevier recorded a later uptake, consistent with the publisher's historical reluctance to engage in negotiations with library consortia (Fraser et al., 2023). However, by 2023, the share of open access enabled by transformative agreements appeared to stabilise for all three publishers (see Figure 5C). Interestingly, the differences between first and corresponding authorships were more pronounced at the publisher level. In Scopus, for example, the share of open access via transformative agreements measured by the corresponding authorship was higher for Springer Nature in 2023 than when using the first authorship.

---

[7] `https://webofscience.help.clarivate.com/en-us/Content/open-access.html`

[8] See Unpaywall version history related to Wiley fixes:

`https://github.com/search?q=repo%3Aourresearch%2Foadoi+wiley&type=commits`

Open access by country. When comparing countries, consistent patterns were observed across data sources for the five-year period from 2019 to 2023. Figure 6 presents hybrid open access indicators by country, comparing hoaddata (x-axis) with Web of Science and Scopus (y-axis). Indicators calculated from these proprietary databases are shown for both the first and corresponding authors, with full counting used to account for multiple country affiliations (Hottenrott et al., 2021).

In terms of article output by country (see Figure 6A), strong positive correlations were observed across data sources and author roles (Spearman rank correlation $\rho > .9$, $p < 0.001$). Between 2019 and 2023, China was the most productive country, followed by the United States and, by a certain margin, India, the United Kingdom, and Germany. The analysis of authorship roles revealed minimal variation, indicating that the first and corresponding authors were typically from the same country.

When examining the percentage of open access articles in hybrid journals (Figure 6B), a different pattern emerged. Authors affiliated with institutions from medium-sized European countries, such as Sweden, the Netherlands, Finland, and Hungary, made a large proportion of their articles open access. Germany and the United Kingdom also had approximately 40% of their output available as open access. In contrast, non-OECD countries showed notably lower adoption of hybrid open access, with South Africa being the only BRICS member well-represented in the data. The United States also demonstrated a relatively low proportion of open access articles. These findings were consistent across all databases. However, France was better represented in Web of Science, likely because of its agreement with Elsevier starting in 2019, which allowed delayed open access under the publisher's user license (Rabesandratana, 2019). This licence was not classified as hybrid open access in either Scopus or hoaddata. In all cases, Spearman rank correlations were $\rho > .9$, $p < 0.001$, indicating a high level of correlation between the databases and authorship roles considered.

Transformative agreements appeared to be a key driver of national open access growth (see Figure 6C and D). OECD members accounted for the majority of open access articles enabled by transformative agreements. As a notable exception, South Africa again featured prominently, as the South African National Library and Information Consortium (SANLiC) successfully negotiated transformative agreements with major publishers from 2022 onward.[9] The results were consistent across data sources. However, Wiley's open access surplus in hoaddata led to better rankings for countries where Wiley played a substantial role, such as Germany, where the DEAL consortium negotiated its first transformative agreements with publishers that started in July 2019.

Medium-sized European countries again showed a high proportion of hybrid open access through transformative agreements (see Figure 6D), highlighting the impact of this licensing model across all three data sources. In contrast, the United States had a low proportion of hybrid open access enabled by agreements, suggesting that a substantial number of open access articles from US-based

---

[9] `https://sanlic.ac.za/read-and-publish-agreements/`
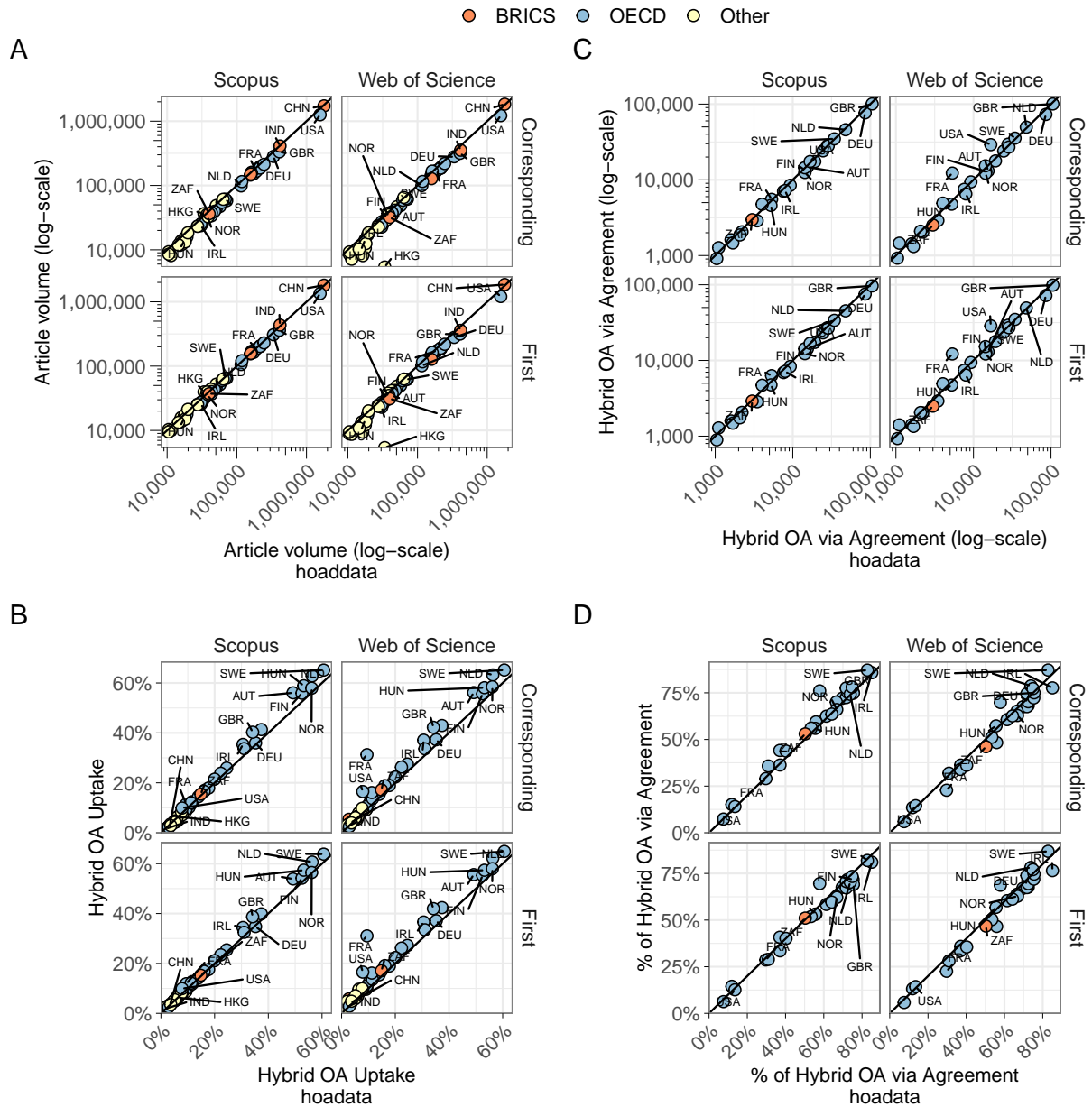
Figure 6. Comparison of hybrid open access by country, 2019-2023. Scatterplots distinguish between proprietary databases Scopus and Web of Science, and author role. The x-axis shows hoaddata indicators. A) Five-years article volume, B) open access percentage in hybrid journals, (A and B limited to countries with > 10.000 articles published), C) number and D) percentage of open access articles enabled by transformative agreements (limited to countries with > 1.000 open access articles enabled by transformative agreements). Line represents line of equality. An interactive version is accessible via: `https://najkoja-hoa-replication.hf.space/`

authors were likely financed through other means.

In all cases, strong positive correlations were observed using Spearman's rank correlation: $\rho > .9$, $p < 0.001$ between data sources and authorship roles, when considering countries with a minimum of 1,000 open access articles enabled by transformative agreements between 2019 and 2023.

When this limitation was removed, the correlation remained strong ($\rho > .85$, $p < 0.001$). This difference may signal countries where only a few institutions had transformative agreements in place, as opposed to those participating in national consortia with broader participation.

4 Discussion

This study of over 13,000 hybrid journals shows a substantial rise in open access due to transformative agreements between 2019 and 2023, although most articles remained paywalled. While transformative agreements accounted for the majority of open access, many articles continue to become open through the payment of individual publication fees. Hybrid open access and transformative agreements remain concentrated among a small group of large commercial publishers, with high-income European countries—alongside South Africa—showing high adoption rates. In contrast, the three most productive countries, China, the United States, and India, show a substantially lower adoption in transformative agreements. Open questions remain as to whether this uneven distribution reflects temporary implementation gaps, inherent inequities in the transformative agreement model, or deliberate avoidance of such agreements.

The findings were consistent across the investigated open data source hoaddata, derived from Crossref and OpenAlex, and the established proprietary bibliometric databases Scopus and Web of Science. This is consistent with previous studies (Akbaritabar et al., 2024; Alperin et al., 2024) and rankings (N. J. van Eck et al., 2024). Overall, the results show strong correlations by country affiliation, which supports the use open metadata for large-scale analyses of hybrid open access. However, the observed differences in journal coverage and metadata availability warrant further discussion, affecting not only open data sources but also proprietary databases when used in isolation.

The coverage analysis revealed that hybrid journals are well indexed in all three data sources, particularly in terms of article coverage, reflecting the dominance of major publishers whose established journal portfolios are comprehensively indexed in proprietary databases (Bellen et al., 2024). Differences emerge for journals targeting practitioners or local non-English language communities, with many such titles indexed exclusively in Crossref and Scopus. Using Crossref as a bibliometric database in hoaddata demonstrated a particular strength in identifying newly established hybrid journals, a notable finding given that transformative agreements primarily target existing subscription-based journals. The landscape of hybrid journal publishing thus differs markedly from that of fully open access journals. Comparing the coverage of OpenAlex, Scopus and Web of Science, Simard et al. (2024) indicate that only half of the fully open access journals listed in the DOAJ are also indexed in Scopus and Web of Science. Notably, journals that charge no publication fees ("diamond journals") are absent from the selective Web of Science, which reinforces existing disparities in the indexing of under-represented research communities and regions in selective bibliometric databases (Simard et al., 2024).

A frequently reported limitation of studying open access with less selective databases is the lack of corresponding authorship information (Fraser et al., 2023; Haucap et al., 2021; Shu & Larivière, 2023). However, this analysis demonstrates that indicators based on first authors, which have often been used as a proxy for determining open access funding, and corresponding authors show a high level of correlation, reflecting disciplinary norms in scholarly publishing with regard to contributorship and author roles and positions (Larivière et al., 2016). The first author typically conduct the main research underlying a paper, while the corresponding author often supervises the research (Fox et al., 2018; Mattsson et al., 2010). Unsurprisingly, measures based on first or corresponding authorships are strongly correlated, suggesting that these authors share the same country affiliation. In most cases, the first author is identical to the corresponding author (Chinchilla-Rodríguez et al., 2024). Despite this correlation, the study observed a recent slight increase in open access articles by corresponding authors over first authors across Springer Nature hybrid journals. This prompts questions about how institutional open access sponsorship practices influence author roles and assignments within co-author teams, especially as funding opportunities vary (Gumpenberger et al., 2018). Previous research has highlighted how institutionalised bibliometric practices can affect the valuation of authorship positions (Helgesson, 2020), suggesting the need to monitor the potential influences of the availability of open access funding on authorship roles (Maddi & Silva, 2024).

Another critical data element in the study is affiliation data, which is essential for estimating open access enabled by transformative agreements. Although OpenAlex's affiliation coverage is less comprehensive, which likely reduced the number of articles confidently attributed to transformative agreements in this study, it still showed high correlations with Scopus and Web of Science at the country level. However, OpenAlex's native ROR-ID integration offers a distinct advantage, allowing a more reliable identification of agreement-enabled articles compared to Scopus and Web of Science, which require reconciliation with proprietary organisation identifiers. Future studies based on Web of Science will benefit from the recent integration of ROR-IDs, announced by the end of 2024.

The database comparison revealed important discrepancies in open access evidence. Crucially, not all publishers share CC licence metadata via Crossref, a limitation that becomes apparent when contrasting Crossref with Unpaywall's data in Scopus and Web of Science. While Unpaywall can detect such gaps by parsing publishers' websites for open access licences, it nevertheless missed a substantial number of CC-licensed open access articles from Wiley journals indexed in Crossref, likely because of parsing errors on the publisher's website. This resulted in fewer open access articles being recorded in Scopus and Web of Science. Further inconsistencies emerged between Scopus and Web of Science, despite both relying on Unpaywall: Web of Science erroneously labels Elsevier's delayed open access as hybrid, whereas Scopus correctly categorises it. Notably, Scopus and Crossref showed greater alignment than that observed in a related comparison between Scopus and OpenAlex (Alperin et al., 2024). This

likely reflects the use of a curated list to identify hybrid open access journals, rather than depending solely on article-level tags. Given these variations, which may compromise comparability, research and monitoring exercises should avoid relying on a single source. Instead, the selection of open access evidence should be cross-verified using multiple sources and snapshots that can be used to track changes in the data over time (Huang et al., 2020). Incorporating expert-curated journal lists can also help prevent misclassification based on business models (Visser et al., 2021).

It is important to note that the study's estimates of articles from institutions involved in transformative agreements are approximations because of a lack of access to invoice data, which is not usually shared by library consortia and publishers. Furthermore, the open data sources for transformative agreements, including the cOAlition S Journal Checker Tool and the ESAC registry, are based on voluntary effort, crowd-sourced from various consortia. However, recent validation using Dutch research information demonstrates the reliability of such an open approach for the assessment of articles under transformative agreements (Jonge et al., 2025). With enhanced open metadata compliance attributable to evolving standards and initiatives in support of negotiations with publishers, particularly through the ESAC initiative and Barcelona Declaration on Open Research Information, the situation is likely to improve. In conclusion, this study has shown that open metadata are well-suited for analysing the transition from subscription-based journal publishing to full open access.

Author contributions

The author confirms the sole responsibility for the following: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Competing interests

The author declares no competing interests.

Funding information

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) under grants 16WIK2301E / 16WIK2101F.

The proprietary bibliometric data from Scopus and Web of Science were provided by the German Competence Network for Bibliometrics under BMBF grant 16WIK2101A.

Data and code availability

Data source code supplement is available via GitHub:
`https://github.com/njahn82/hoa_validation`. An interactive dashboard, which allows the browsing of metrics by country and publisher, is available on HuggingFace Spaces:
`https://huggingface.co/spaces/najkoja/hoa_replication`.

Please note that according to the licence terms for Scopus and Web of Science, as provided by

the German Competence Network for Bibliometrics, only aggregated data accompanying the scientific publication can be made publicly available for replication purposes.

Open metadata provided by Crossref, OpenAlex, and cOAlition S Journal Checker Transformative Agreement Data are freely available for re-use.

References

Akbaritabar, A., Theile, T., & Zagheni, E. (2024). Bilateral flows and rates of international migration of scholars for 210 countries for the period 1998-2020. *Scientific Data*, *11*(1). https://doi.org/10.1038/s41597-024-03655-9

Alperin, J. P., Portenoy, J., Demes, K., Larivière, V., & Haustein, S. (2024). *An analysis of the suitability of OpenAlex for bibliometric analyses.* arXiv. https://arxiv.org/abs/2404.17663

Aspesi, C., & Brand, A. (2020). In pursuit of open science, open access is not enough. *Science*, *368*(6491), 574–577. https://doi.org/10.1126/science.aba3763

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Babini, D., Chan, L., Hagemann, M., Joseph, H., Kuchma, I., & Suber, P. (2022). *The Budapest Open Access Initiative-20th. Anniversary recommendations (BOAI20).* https://www.budapestopenaccessinitiative.org/boai20/

Bellen, S. van, Alperin, J. P., & Larivière, V. (2024). *The oligopoly of academic publishers persists in exclusive database.* arXiv. https://arxiv.org/abs/2406.17893

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Björk, B.-C. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, *63*(8), 1496–1504. https://doi.org/10.1002/asi.22709

Borrego, Á., Anglada, L., & Abadal, E. (2021). Transformative agreements: Do they pave the way to open access? *Learned Publishing*, *34*(2), 216–232. https://doi.org/10.1002/leap.1347

Brayman, K., Devenney, A., Dobson, H., Marques, M., & Vernon, A. (2024). *A review of transitional agreements in the UK.* Zenodo. https://doi.org/10.5281/zenodo.10787392

Butler, L.-A., Matthias, L., Simard, M.-A., Mongeon, P., & Haustein, S. (2023). The oligopoly's shift to open access: How the big five academic publishers profit from article processing charges. *Quantitative Science Studies*, *4*(4), 778–799. https://doi.org/10.1162/qss_a_00272

Campbell, C., Dér, Á., Geschuhn, K., & Valente, A. (2022). How are transformative agreements transforming libraries? *87th IFLA World Library and Information Congress (WLIC) / 2022 in*

Dublin, Ireland. `https://repository.ifla.org/handle/123456789/1973`

Céspedes, L., Kozlowski, D., Pradier, C., SainteMarie, M. H., Shokida, N. S., Benz, P., Poitras, C., Ninkov, A. B., Ebrahimy, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2025). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology.* `https://doi.org/10.1002/asi.24979`

Chinchilla-Rodríguez, Z., Costas, R., Robinson-García, N., & Larivière, V. (2024). Examining the quality of the corresponding authorship field in Web of Science and Scopus. *Quantitative Science Studies*, *5*(1), 76–97. `https://doi.org/10.1162/qss_a_00288`

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics.* `https://doi.org/10.1007/s11192-025-05293-3`

Dér, Á. (2025). What gets missed in the discourse on transformative agreements. *Katina Magazine.* `https://doi.org/10.1146/katina-20250212-1`

Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics*, *113*(1), 219–236. `https://doi.org/10.1007/s11192-017-2483-y`

Eck, N. J. van, & Waltman, L. (2022). *Crossref as a source of open bibliographic metadata.* MetaArXiv. `https://doi.org/10.31222/osf.io/smxe5`

Eck, N. J. van, Waltman, L., & Neijssel, M. (2024). *Launch of the CWTS Leiden Ranking Open Edition 2024.* Leiden Madtrics. `https://doi.org/10.59350/r512t-r8h93`

Fox, C. W., Ritchey, J. P., & Paine, C. E. T. (2018). Patterns of authorship in ecology and evolution: First, last, and corresponding authorship vary with gender and geography. *Ecology and Evolution*, *8*(23), 11492–11507. `https://doi.org/10.1002/ece3.4584`

Fraser, N., Hobert, A., Jahn, N., Mayr, P., & Peters, I. (2023). No deal: German researchers' publishing and citing behaviors after big deal negotiations with Elsevier. *Quantitative Science Studies*, *4*(2), 325–352. `https://doi.org/10.1162/qss_a_00255`

Geschuhn, K., & Stone, G. (2017). It's the workflows, stupid! What is required to make "offsetting" work for the open access transition. *Insights the UKSG Journal*, *30*(3), 103–114. `https://doi.org/10.1629/uksg.391`

Gumpenberger, C., Hölbling, L., & Gorraiz, J. I. (2018). On the issues of a "corresponding author" field-based monitoring approach for gold open access publications and derivative cost calculations. *Frontiers in Research Metrics and Analytics*, *3.* `https://doi.org/10.3389/frma.2018.00001`

Harrell Jr, F. E. (2003). Hmisc: Harrell miscellaneous. In *CRAN: Contributed Packages.* The R Foundation. `https://doi.org/10.32614/cran.package.hmisc`

Haucap, J., Moshgbar, N., & Schmal, W. B. (2021). The impact of the German "DEAL" on competition in the academic publishing market. *Managerial and Decision Economics*, *42*(8),

2027–2049. https://doi.org/10.1002/mde.3493

Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2024). *Analysis of the publication and document types in OpenAlex, web of science, scopus, pubmed and semantic scholar.* arXiv. https://doi.org/10.48550/ARXIV.2406.15154

Haustein, S., Schares, E., Alperin, J. P., Hare, M., Butler, L.-A., & Schönfelder, N. (2024). *Estimating global article processing charges paid to six publishers for open access between 2019 and 2023.* https://arxiv.org/abs/2407.16551

Helgesson, G. (2020). Authorship order and effects of changing bibliometrics practices. *Research Ethics*, *16*(1–2), 1–7. https://doi.org/10.1177/1747016119898403

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Holden, L., Skoie, M., Røeggen, V., Bjerde, K. W., Wenaas, L., Bakke, P., Løvhaug, J. W., Karlsen, E. S., & Qvenild, M. (2023). *Strategi for vitenskapelig publisering etter 2024.* Sikt. https://doi.org/10.18711/2KZ1-BA97

Hottenrott, H., Rose, M. E., & Lawson, C. (2021). The rise of multiple institutional affiliations in academia. *Journal of the Association for Information Science and Technology*, *72*(8), 1039–1058. https://doi.org/10.1002/asi.24472

Huang, C.-K., Neylon, C., Hosking, R., Montgomery, L., Wilson, K., Ozaygen, A., & Brookes-Kenworthy, C. (2020). *Evaluating institutional open access performance: Sensitivity analysis.* bioRxiv. https://doi.org/10.1101/2020.03.19.998542

Jahn, N. (2024). *hoaddata: Data about hybrid open access journal publishing (v.0.3).* https://github.com/subugoe/hoaddata/releases/tag/v.0.3

Jahn, N. (2025). How open are hybrid journals included in transformative agreements? *Quantitative Science Studies*, *6*, 242–262. https://doi.org/10.1162/qss_a_00348

Jahn, N., Haupka, N., & Hobert, A. (2023). *Analysing and reclassifying open access information in OpenAlex.* Scholarly Communication Analytics Blog. https://subugoe.github.io/scholcomm_analytics/posts/oalex_oa_status/

Jahn, N., Matthias, L., & Laakso, M. (2022). Toward transparency of hybrid open access through publisher-provided metadata: An article-level study of Elsevier. *Journal of the Association for Information Science and Technology*, *73*(1), 104–118. https://doi.org/10.1002/asi.24549

Jonge, H. de, Kramer, B., & Sondervan, J. (2025). *Tracking transformative agreements through open metadata: Method and validation using Dutch Research Council NWO funded papers.* MetaArXiv. https://doi.org/10.31222/osf.io/tz6be_v1

Kiley, R. (2024). *Transformative Journals: Analysis from the 2023 reports.* cOAlition S. https:

//www.coalition-s.org/blog/transformative-journals-analysis-from-the-2023-reports/

Kohls, A., & Mele, S. (2018). Converting the literature of a scientific field to open access through global collaboration: The experience of SCOAP3 in particle physics. *Publications*, *6*(2), 15. https://doi.org/10.3390/publications6020015

Krassowski, M. (2020). *ComplexUpset*. Zenodo. https://doi.org/10.5281/zenodo.3700590

Laakso, M., & Björk, B.-C. (2016). Hybrid open access–a longitudinal study. *Journal of Informetrics*, *10*(4), 919–932. https://doi.org/10.1016/j.joi.2016.08.002

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, *46*(3), 417–435. https://doi.org/10.1177/0306312716650046

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992. https://doi.org/10.1109/tvcg.2014.2346248

Maddi, A., & Silva, J. A. T. da. (2024). Beyond authorship: Analyzing contributions in PLOS ONE and the challenges of appropriate attribution. *Journal of Data and Information Science*, *9*(3), 88–115. https://doi.org/10.2478/jdis-2024-0015

Maisano, D. A., Mastrogiacomo, L., Ferrara, L., & Franceschini, F. (2025). A large-scale semi-automated approach for assessing document-type classification errors in bibliometric databases. *Scientometrics*. https://doi.org/10.1007/s11192-025-05244-y

Marwick, B., Boettiger, C., & Mullen, L. (2018). Packaging data analytical work reproducibly using R (and friends). *The American Statistician*, *72*(1), 80–88. https://doi.org/10.1080/00031305.2017.1375986

Mattsson, P., Sundberg, C. J., & Laget, P. (2010). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, *87*(1), 99–105. https://doi.org/10.1007/s11192-010-0310-9

McCabe, M. J., & Mueller-Langer, F. (2024). Open access is shaping scientific communication. *Science*, *385*(6714), 1170–1172. https://doi.org/10.1126/science.adp8882

Melero-Fuentes, D., Aguilar-Moya, R., Valderrama-Zurián, J.-C., & Gorraiz, J. (2025). Evolution and effect of meeting abstracts in JCR journals. *Journal of Informetrics*, *19*(1), 101631. https://doi.org/10.1016/j.joi.2024.101631

Momeni, F., Mayr, P., Fraser, N., & Peters, I. (2021). What happens when a journal converts to open access? A bibliometric analysis. *Scientometrics*, *126*(12), 9811–9827. https://doi.org/10.1007/s11192-021-03972-5

Muñoz-Vélez, H., Pallares, C., Echavarría, A. F., Contreras, J., Pavas, A., Bello, D., Rendón, C., Calderón-Rojas, J., & Garzón, F. (2024). Strategies for negotiating and signing transformative

agreements in the Global South: The Colombia Consortium experience. *Journal of Library Administration*, *64*(1), 80–98. https://doi.org/10.1080/01930826.2023.2287945

Pinfield, S. (2015). Making open access work: The "state-of-the-art" in providing open access to scholarly literature. *Online Information Review*, *39*(5), 604–636. https://doi.org/10.1108/oir-05-2015-0167

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts.* https://arxiv.org/abs/2205.01833

R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rabesandratana, T. (2019). Elsevier deal with France disappoints open-access advocates. *Science.* https://doi.org/10.1126/science.aba5656

Robinson-Garcia, N., Costas, R., & Leeuwen, T. N. van. (2020). Open access uptake by universities worldwide. *PeerJ*, *8*, e9410. https://doi.org/10.7717/peerj.9410

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: A scoping review. *Royal Society Open Science*, *9*(1). https://doi.org/10.1098/rsos.211032

Rothfritz, L., Schmal, W. B., & Herb, U. (2024). *Trapped in transformative agreements? A multifaceted analysis of >1,000 contracts.* arXiv. https://arxiv.org/abs/2409.20224

Schimmer, R., Geschuhn, K., & Vogler, A. (2015). *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access.* Max Planck Digital Library. https://doi.org/10.17617/1.3

Schmal, W. B. (2024). *La révolution dévore ses enfants: Pricing implications of transformative agreements.* arXiv. https://doi.org/10.48550/ARXIV.2403.03597

Schmidt, M., Rimmert, C., Stephen, D., Lenke, C., Donner, P., Gärtner, S., Taubert, N., Bausenwein, T., & Stahlschmidt, S. (2024). *The data infrastructure of the German Kompetenznetzwerk Bibliometrie: An enabling intermediary between raw data and analysis.* Zenodo. https://doi.org/10.5281/zenodo.13935407

Shu, F., & Larivière, V. (2023). The oligopoly of open access publishing. *Scientometrics*, *129*(1), 519–536. https://doi.org/10.1007/s11192-023-04876-2

Simard, M.-A., Basson, I., Hare, M., Lariviere, V., & Mongeon, P. (2024). *The open access coverage of OpenAlex, Scopus and Web of Science.* arXiv. https://doi.org/10.48550/ARXIV.2404.01985

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of

Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. `https://doi.org/10.1007/s11192-021-03948-5`

Stahlschmidt, S., & Stephen, D. (2022). From indexation policies through citation networks to normalized citation impacts: Web of Science, Scopus, and Dimensions as varying resonance chambers. *Scientometrics*, *127*(5), 2413–2431. `https://doi.org/10.1007/s11192-022-04309-6`

Suber, P. (2012). *Open access.* The MIT Press. `https://doi.org/10.7551/mitpress/9286.001.0001`

Visser, M., Eck, N. J. van, & 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. `https://doi.org/10.1162/qss_a_00112`

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. `https://doi.org/10.21105/joss.01686`

Widding, A. S. (2024). Beyond transformative agreements: Ways forward for universities. *European Review*, *32*(S1), S28–S38. `https://doi.org/10.1017/s1062798724000036`

Zhang, L., Cao, Z., Shang, Y., Sivertsen, G., & Huang, Y. (2024). Missing institutions in OpenAlex: Possible reasons, implications, and solutions. *Scientometrics*, *129*(10), 5869–5891. `https://doi.org/10.1007/s11192-023-04923-y`