

Reassessment of the agreement in retraction indexing across 4 multidisciplinary sources: Crossref, Retraction Watch, Scopus, and Web of Science

Malik Oyewale Salami*, Corinne McCumber*, and Jodi Schneider*†
* School of Information Sciences, University of Illinois Urbana-Champaign, USA
† Harvard Radcliffe Institute for Advanced Study, USA

Keywords: student submission, retraction status, data quality, indexing, retraction indexing

BACKGROUND

As of April 2023, over 1 in 5 DOIs had discrepancies in retraction indexing among the 49,924 DOIs indexed as retracted in at least one of Crossref, Retraction Watch, Scopus, and Web of Science (Schneider et al., 2023). Since then, in September 2023, the Retraction Watch database was CC0-licensed (Hendricks et al., 2023), and over 10,000 publications were retracted in 2023 (Van Noorden, 2023), mostly late in the year. Here, we determine what changed in 15 months.

METHODS AND DATA

To compare with the April 2023 union list (Schneider et al., 2023), we constructed a union list in July 2024 following the methodology of Schneider et al. (2023). This is part of a larger project in preparation (Salami et al., 2024). For this extended abstract, code (Salami & McCumber, 2024b) and redacted data (Salami & McCumber, 2024a) are available. 83,771 unique DOIs were indexed as retracted publications by one or more of Crossref, Retraction Watch, Scopus, and Web of Science as shown in Table 1.

Table 1. Publications indexed as retracted from multidisciplinary sources

Source	Query results	Search Date	Records with DOIs	Records without DOIs	Duplicate DOIs removed
Crossref	30,921	2024-07-03	30,918	0	3
Retraction Watch	53,800	2024-07-03	45,419	5,922	2,459
Scopus	31,268	2024-07-05	30,878	338	52
Web of Science	32,772	2024-07-03	31,683	1,062	27
Total	148,761		138,898	7,322	2,541
Total Unique DOIs	83,771				

RESULTS

Changes in the union lists over time

We identified 34,680 additional items with DOIs, a 69.5% increase in the 15-month period. Changes between the April 2023 and July 2024 union lists are summarized in Figure 1, which shows newly added and missing DOIs from January-April 2023; May-December 2023; and January-July 2024, as well as differences in records with no DOIs that were dropped.

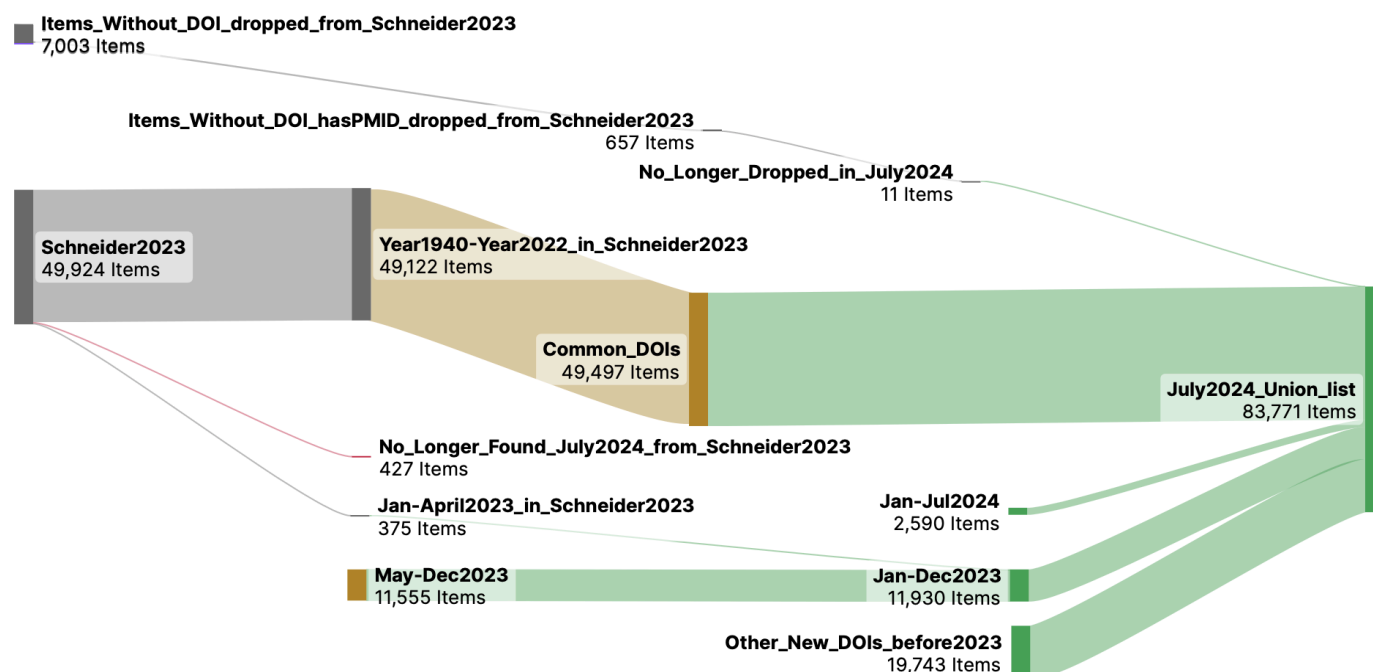


Figure 1. Changes from April 2023 to July 2024 in a Sankey diagram (created using sankeyart.com)

Retraction indexing agreement over time

More than twice the number of items retracted in all of 2023 were added to the July 2024 union list: 34,680 DOIs were newly added, and of these 34,680 items, 641 items had PMIDs but no recorded DOIs. Retraction indexing agreement by source dropped dramatically from 92.9% to 37.4% (30,918/82,732 covered) for Crossref; from 68.3% to 45.7% (30,878/67,526 covered) for Scopus, and from 95.7% to 62.6% (31,683/50,598 covered) for Web of Science Core. Retraction indexing agreement was not calculated for Retraction Watch because it only covers items it deems retracted publications. Figure 2 shows the distribution of retraction indexing agreement scores by item.

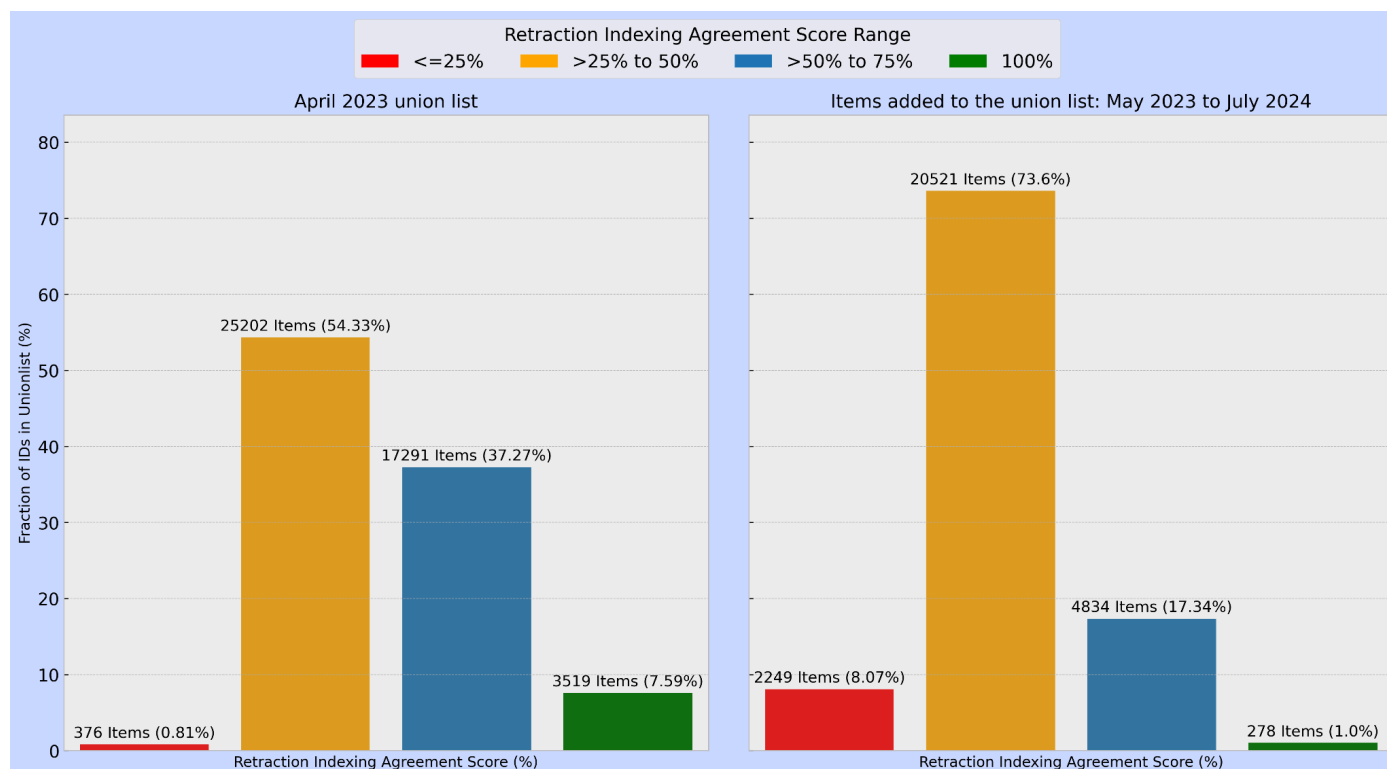


Figure 2. Distribution of retraction indexing agreement scores of items (excluding 9,501 items that are only indexed and covered in a single source): April 2023 (left) and new items from the July 2024 union list (right).

Based on coverage checks July 30, 2024 (Web of Science); August 2, 2024 (Scopus); and August 6, 2024 (Crossref).

Investigating DOIs that no longer appeared in the July 2024 union list

To investigate the 427 DOIs no longer found in the July 2024 union list, we manually examined a sample of 77 DOIs. For some DOIs, none of the sources still index them as retracted; other DOIs were no longer covered in any of the 4 sources. The changes are mixed: we found both data quality improvements (covered DOIs that are not retracted¹ according to the publisher website) as well as data quality reductions (covered DOIs that are retracted according to the publisher).

DISCUSSION AND CONCLUSIONS

We investigated discrepancies in retraction indexing in 4 multidisciplinary databases and found even worse retraction indexing agreement than our prior work (Schneider et al., 2023) despite hugely increased coverage. This indicates massive disagreement among these 4 sources over the span of 15 months (April 2023 to July 2024). One possible reason is that some databases have increased coverage of items indexed as retracted in other sources. For instance, between February 2 and February 13, 2024, we observed a jump from 14,920 to 27,493 items indexed as retracted in Web of Science Core, which they stated was due to a new data load. Data quality improvements may have partly come from Crossref's recent acquisition and CC0-licensing of the Retraction Watch dataset (Hendricks et al., 2023). Yet, potential issues in database indexing processes could include delayed updates, errors in retraction indexing, and inconsistencies in data integration. Our results highlight the need for continued refinement of retraction indexing practices.

DATA AVAILABILITY

Among our sources, data is openly available only for Crossref and Retraction Watch (Salami & McCumber 2024a). Other data used for this study is licensed by each source and was retrieved using licensed APIs with privileged API keys.

ACKNOWLEDGEMENTS

Alfred P. Sloan Foundation G-2022-19409 Reducing the Inadvertent Spread of Retracted Science II: Research and Development towards the Communication of Retractions, Removals, and Expressions of Concern. U.S. National Science Foundation 2046454 CAREER: Using network analysis to assess confidence in research synthesis. Jodi Schneider was partly funded by the 2024-2025 Perrin Moorhead Grayson and Bruns Grayson Fellowship at the Harvard Radcliffe Institute for Advanced Study. Thanks to Alick Bird of Elsevier, Ivan Oransky of Retraction Watch, Martyn Rittman of Crossref, and Tilla Edmunds of Clarivate for their thoughts regarding changes in the retraction indexing landscape. Thanks to Christoph Stade, Heng Zheng, and Jess Rogovin for feedback on a draft.

CRedit

Corinne McCumber: Software, Validation, Writing – original draft, Writing – review & editing

Jodi Schneider: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Malik Salami: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

REFERENCES

- Hendricks, G., Lammey, R., Lucy, O., Bilder, G., & Pentz, E. (2023, September 12). News: Crossref and Retraction Watch Crossref acquires Retraction Watch data and opens it for the scientific community. *Crossref*. <https://doi.org/10.13003/c23rw1d9>
- Salami, M. O. & McCumber, C. (2024a). *Dataset for Reassessment of the agreement in retraction indexing across 4 multidisciplinary sources: Crossref, Retraction Watch, Scopus, and Web of Science* (Version 1) [Data set]. University of Illinois Urbana-Champaign Databank. https://doi.org/10.13012/B2IDB-8457537_V1
- Salami, M. O. & McCumber, C. (2024b). Retraction indexing agreement: ASIS&T MET-STI2024 Paper Final Code (1.2.1) [Python]. Zenodo. <https://doi.org/10.5281/zenodo.14052823>
- Salami, M. O., McCumber, C., & Schneider, J. (2024). *Analyzing the consistency of retraction indexing*. MetaArXiv. <https://doi.org/10.31222/osf.io/gvfk5>
- Schneider, J., Lee, J., Zheng, H., & Salami, M. O. (2023). *Assessing the agreement in retraction indexing across 4 multidisciplinary sources: Crossref, Retraction Watch, Scopus, and Web of Science*. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023). <https://doi.org/10.55835/6441e5cae04dbe5586d06a5f>
- Van Noorden, R. (2023). More than 10,000 research papers were retracted in 2023—A new record. *Nature*, 624(7992), 479–481. <https://doi.org/10.1038/d41586-023-03974-8>

¹ This includes the 4/425 DOIs manually checked in (Schneider et al., 2023), all of which we considered not retracted.