

A Survey of Exclusively Data Journals and How They Are Indexed by Scientific Databases

Kai Li

Renmin University of China, China
kai.li@ruc.edu.cn

Chuyi Lu

Library of Northwest Agriculture &
Forestry University, China
luchuyi@nwfau.edu.cn

Chenyue Jiao

University of Illinois at Urbana-
Champaign, USA
cjiao4@illinois.edu

ABSTRACT

As data becomes omnipresent in the scientific system, a new academic genre aiming to describe data objects (*data papers*) and the venue to publish these articles (*data journals*) gradually emerged from the end of the 2000s. However, it is largely unknown how much these scientific outputs are indexed in scientific databases, which has greatly prevented them from being thoroughly studied in large-scale, quantitative studies. This poster presents our preliminary efforts to address this gap, by compiling a list of data journals that primarily accept data papers (i.e., *exclusively data journals*) and examining their presence in four major scientific databases. Our results indicate that exclusively data journals are comprehensively indexed in Crossref and Dimensions, two relatively new scientific databases, which can be used to conduct future studies on data papers and journals. The next steps of our project are also discussed in this poster.

KEYWORDS

Data papers; Data journals; Scientific database; Quantitative science studies.

INTRODUCTION

Data publication has become an important framework to make data more transparent and accessible in scientific communities (Kratz & Strasser, 2014). A key implementation of this framework is to publish datasets into academic papers (i.e., *data papers*) that aim to describe data objects, in order to facilitate their sharing and reuse (Chavan & Penev, 2011). As a distinct academic genre from regular research articles, data papers have been increasingly investigated as a research object during the past few years (Kim, 2020; Li et al., 2020; Thelwall, 2020).

Since the late 2000s, more *data journals* have been created as the venue for the publication of data papers. Some of these journals have high impacts and popularity among researchers, such as *Scientific Data* and *Earth System Science Data* (Jiao & Darch, 2020), even though there is a strong variance across journals. Candela and colleagues offered a highly-cited survey of data journals that have been founded by the beginning of 2014, where they found seven *exclusively data journals* (i.e., journals that ONLY publish data papers) and over 100 *mixed data journals* (Candela et al., 2015). However, no other survey has been conducted on this rapidly developing topic since then to evaluate whether newer data journals have been created during subsequent years and particularly how these journals are covered by scientific databases. This has prevented data papers from being analyzed in large-scale studies.

This project aims to fill the gap by compiling a list of exclusively data journals and examining how these journals are indexed in the most popular scientific databases, including the Web of Science (WoS), Scopus, Crossref and Dimensions, to facilitate quantitative studies on this topic in the future. As the first step of this project, this poster reports methods and preliminary results of how we identified major exclusively data journals and queried their publications in the four databases. In the end of the poster, we also discussed the next steps of this project.

METHOD

We used the following resources to identify **data journals**: (1) Candela and colleagues' survey (Candela et al., 2015), (2) data journal lists created by academic libraries and other parties (One example is the list of data journals created by the University of Edinburgh: <https://www.wiki.ed.ac.uk/display/datasources+of+dataset+peer+review>) indexed by Google, (3) journals with "data" or "database" in the title included in the Journal Citation Reports or Scopus List of Journals, and (4) manual examination of whether the journal is a data journal or a regular research journal. From these sources, we further selected **exclusively data journals** based on the following criteria: (1) the journal primarily accepts data papers based on its scope, (2) the journal is active as of January 2021, and (3) the journal only publishes English-language articles.

The above steps were conducted from January to May 2021, and we identified 25 exclusively data journals that meet all criteria. For all these journals, we gathered information from their websites, especially when the journals were founded. Moreover, we also collected the numbers of publications being indexed in the Web of Science, Scopus, Crossref and Dimensions. For the latter task, we considered all publications from the 25 journals that are (1) in

84th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 3, 2021 | Salt Lake City, UT. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

English-language, (2) published by the end of 2020, and (3) being classified into either research articles or data papers in these four databases.

RESULTS

The list of all 25 exclusively data journals identified from this research has been deposited on Figshare (Li, 2021). Figure 1 summarizes the years in which these journals were established. We preliminarily separated the history of data journals into the three periods. While three journals were established before 2000 (all before 1975), none of them was covered by Candela's survey. Even though all these journals (most notably *Nuclear Data Sheets*) meet our criteria, we acknowledge that their publications may not be totally consistent with how data papers are defined today. More recently, the year of 2014 is also the year where six data journals were established, including *Scientific Data* and *Data in Brief*, two of the most important data journals.

Period	Journals
Before 2000	3
2000-2013	8
2014-2016	14

Table 1. Number of data journals established over time

Table 2 illustrates the number of journals and articles covered by the four databases. As shown in the table, the WoS and Scopus have rather limited coverage of these data journals, echoing existing evidence about how these databases index non-traditional materials, such as digital humanities journals (Spinaci et al., 2020). On the other side of the spectrum, however, Crossref and Dimensions have similar numbers of journals and publications from our journal list. This indicates that both of the latter databases may have strong potentials for future studies on data papers, despite the fact that more efforts are needed to understand differences between them, especially whether different publication types exist under their relatively coarse metadata.

Database	Journals	Publications
Web of Science	10	16,781
Scopus	16	21,046
Crossref	24	26,913
Dimensions	25	28,042

Table 2. Number of journals and publications from the journal list indexed in the four databases

CONCLUSION

This poster presents our methods to identify exclusively data journals as well as some preliminary results about how these 25 journals are indexed in the four major scientific databases that are frequently used in quantitative science studies. This work is the first step of establishing the infrastructure to support large-scale empirical studies on data papers. Our results show that more exclusively data journals have been founded since 2014, many of which are not covered by the previous survey. Moreover, we also found that both Crossref and Dimensions may serve as solid data sources for this line of research, given their comprehensive coverage of exclusively data journals. As compared, the Web of Science and Scopus, two most popular scientific databases, only cover around half of data journals in our list.

In the next step of this project, we plan to analyze how different publication types (especially data papers and research articles) from these journals are indexed in the databases over time. More importantly, we also plan to include mixed data journals into our analysis to more systematically understand the landscape of data publication. These efforts will contribute to future studies on the roles played by research data and data papers in the scholarly communication system.

REFERENCES

- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747–1762. <https://doi.org/10.1002/asi.23358>
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12(15), 1.

- Jiao, C., & Darch, P. T. (2020). The role of the data paper in scholarly communication. *Proceedings of the Association for Information Science and Technology*, 57(1), e316.
- Kim, J. (2020). An analysis of data paper templates and guidelines: types of contextual information described by data journals. *Science Editing*, 7(1), 16–23.
- Kratz, J. E., & Strasser, C. (2014). Data publication consensus and controversies. *F1000Research*, 3, 1–21. <https://doi.org/10.12688/f1000research.3979.3>
- Li, K. (2021). *List of Exclusively Data Journals*. <https://doi.org/10.6084/m9.figshare.14671599.v1>
- Li, K., Greenberg, J., & Dunic, J. (2020). Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. *Journal of the Association for Information Science and Technology*, 71(2), 172–182.
- Spinaci, G., Peroni, S., & Colavizza, G. (2020). Preliminary Results on Mapping Digital Humanities Research. *L'Associazione per l'Informatica Umanistica e La Cultura Digitale 2020*.
- Thelwall, M. (2020). Data in Brief: Can a mega-journal for data be useful? *Scientometrics*, 124(1), 697–709.