

Research on the Formulation and Implementation Models of Data Policies for Scientific Journals in China: Postprint

Authors: Shen Yan

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose / Significance] The integrated development of data publishing and scientific journals constitutes a “win-win” pathway. Addressing current issues such as the limited volume of data publishing associated with journal articles and the insufficiently stimulated awareness and enthusiasm among authors regarding data publishing, this study investigates the practical experiences of international publishers’ data policies, and proposes mechanisms and policies for Chinese scientific journals to participate in data publishing, thereby providing robust support for advancing data publishing in China and constructing world-class journals. [Method / Process] Through literature review, case analysis, website information collection, and other methods, this paper analyzes the current status of data and article associated publishing in China, as well as the data policies and operational mechanisms of five major international publishers: Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis. [Results / Conclusion] The five publishers implement fundamentally consistent models for data policies: overall policy statements + voluntary adoption by individual journals. Specifically, 68%, 99.9%, 92.2%, 47.0%, and 69.3% of journals under Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis, respectively, explicitly stipulate data policies, primarily in an encouraging manner. China’s publishing community can draw on the practical experiences of international publishers, combine them with the characteristics of Chinese scientific journals, formulate standard norms and implementation models for scientific journal data policies, and gradually promote the integrated development of data publishing and scientific journals in both breadth and depth, moving from few to many and from point to surface.

Full Text

Data Policy Formulation and Implementation Models for Chinese Sci-tech Journals

Shen Yan

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract:

[Purpose/Significance] The integrative development of data publication and sci-tech journals represents a mutually beneficial pathway. However, the current volume of data publications associated with journal articles remains low in China, and authors' awareness and enthusiasm for data publication have yet to be fully stimulated. This study investigates the practical experiences of international publishers' data policies and proposes mechanisms and policies for Chinese sci-tech journals to engage in data publication, providing robust support for advancing data publication and constructing world-class journals in China. **[Methods/Process]** Through literature review, case analysis, and web-based information collection, this paper examines the current state of data-article linked publication in China and analyzes the data policies and operational mechanisms of five major international publishers: Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis. **[Results/Conclusion]** The five publishers employ a consistent implementation model: a general policy statement combined with journal-level autonomy. Among their portfolios, 68%, 99.9%, 92.2%, 47.0%, and 69.3% of journals under Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis respectively have explicitly formulated data policies, though most remain encouragement-based. China's publishing community can draw lessons from these international practices and, in conjunction with the characteristics of Chinese sci-tech journals, establish standardized norms and implementation models for journal data policies, gradually promoting the breadth and depth of integrative development between data publication and sci-tech journals.

Keywords: sci-tech journal; data publication; data policy; operation mechanism; integrative development

1. Introduction

Scientific data, also known as research data, constitute a crucial national strategic resource. With the rise and development of the fourth research paradigm driven by data, scientific data sharing has become particularly necessary and urgent as a key means for data circulation and reuse. Data publication represents an effective approach to promoting data sharing. Data publication can be broadly categorized into three models: (1) independent data publication, where data are stored separately in public data repositories; (2) data as supplementary

materials, published online alongside papers as attachments or appendices; and (3) linked publication, where entity data are stored in journal-designated repositories and linked with scientific discovery papers or data papers to create an interoperable scientific environment. The linked publication of data and papers enables direct access to relevant data when reading articles, thereby facilitating the reuse of scientific data. Conversely, publishing datasets concurrently with (or prior to) papers can reduce research integrity issues such as data fabrication, improve data quality, and ensure the accuracy and validity of research conclusions, thereby enhancing the quality of both papers and journals. Therefore, data publication and journal development can be mutually reinforcing and synergistic. However, in terms of both quantity and degree of association, data publication and sci-tech journals have not yet achieved truly integrative development (characterized by mutual integration, promotion, and common progress).

China has already undertaken some explorations in the integrative development of data publication and sci-tech journals. In June 2014, the Global Change Research Data Publishing & Repository (GCdataPR), supervised by the Chinese Academy of Sciences and co-sponsored by the Institute of Geographic Sciences and Natural Resources Research and the Geographical Society of China, was officially launched, providing a platform for the linked publication of entity data and data papers. In March 2016, GCdataPR and sci-tech journals in the fields of global change and geosciences signed a “Joint Initiative on Linked Publication of Original Data for Papers and Data Sharing,” aiming to promote the linked publication of data for journal articles and facilitate the integrative development of data publication and sci-tech journals. Currently, over 60 domestic and international sci-tech journals have joined this initiative. China has also witnessed the emergence of data journals such as *China Scientific Data*, *Geology in China* (Supplement), *Journal of Agricultural Big Data*, and *Big Earth Data*, which primarily publish data papers. Additionally, some traditional academic journals including *Library Journal* and *Journal of Remote Sensing* have begun publishing data papers. This has created a favorable atmosphere for the integrative development of data publication and journals. Nevertheless, the actual volume of entity dataset publications associated with papers remains unsatisfactory. According to the GCdataPR website, between 2016 and 2020, 65 journals published paper-linked data through GCdataPR, yet the proportion of such publications accounted for less than 2% of papers containing data. This indicates that authors’ awareness and enthusiasm for data publication have not been sufficiently stimulated, and the depth and breadth of integrative development between data publication and sci-tech journals require further enhancement.

Journal data policies refer to policies regarding data associated with published papers, primarily encompassing: descriptions of the benefits of open data sharing; specifications on data types, formats, and related instructions (such as data citation); data quantity limits; preservation methods for different data types; data accessibility and access methods; timelines for authors to provide data; whether data are publicly available, freely accessible, or subject to other restrictions; conditions for data reuse; whether exceptions to the policy are permitted;

and methods for compliance and control.

China currently has 4,477 Chinese sci-tech journals, with an average annual article volume of 324 papers per journal. A survey of selected geography and resource journals revealed that the proportion of data-containing papers ranges from 29.4% to 100.0%, averaging 75.7% (see). Based on this average, over one million entity datasets associated with papers would be generated annually. Journal data policies directly influence the publication of this portion of datasets. If authors understand data policies at the time of submission, the likelihood of data publication would increase substantially. Many international publishers and journals have already established data policies that encourage or mandate authors to deposit paper-linked data in public repositories and provide interconnection services between data and papers. Cases of authors publishing data due to international journal requirements are becoming increasingly common. However, few Chinese journals have formulated explicit data policies. Therefore, China can learn from international publishers' experiences to construct localized data policies and implementation models—determining what types of data policies would suit the characteristics of Chinese sci-tech journals and what measures and mechanisms relevant stakeholders (institutions) should adopt to ensure effective policy implementation.

Domestic scholars have conducted multi-level analyses of international publishers' and journals' data policies for various research purposes. Liu Ying et al. analyzed data publication policies of international academic journals in the medical field from perspectives such as data types, storage locations, and data availability statements. Chen Quanping examined the policy-making bodies and content of data policies using typical international journals as examples. Kong Lihua et al. also analyzed data policies of some international publishers when studying trends, mechanisms, and challenges in data publication. These studies provide theoretical references for domestic journal data policy research, yet comprehensive investigations from a practical perspective that examine international publishers' and journals' data policies and propose operational models for Chinese sci-tech journals remain absent. Therefore, this study selects five major publishers—Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis—and investigates their data policies and operational mechanisms through comprehensive web-based surveys. Based on these findings, we propose data policies and operational models for Chinese sci-tech journals to promote the integrative development of data publication and sci-tech journals in China.

2019 Data-Containing Papers in Selected Geography and Resource Journals

2. Analysis of International Publishers' Data Policies

2.1 Operation Mode

The five publishers employ a consistent implementation model: a general data policy statement combined with journal-level autonomy. Each publisher dedicates a specific section on its official website to introduce its overall data policy,

including definitions of scientific data, the role of data sharing, and methods and levels of data sharing. Individual journals within each portfolio can select appropriate data policies based on their specific circumstances and, when necessary, link to the publisher's website for authors to gain comprehensive understanding. Taking Elsevier's journal *Examples and Counterexamples* as an example (see [Figure 1: see original paper]), the author guidelines include a separate "research data" module that articulates the journal's stance on data sharing—"encourages and enables"—meaning the journal encourages authors to share data and supports linking scientific data with journal articles through DOIs, public repository accession numbers, database identifiers, and other means. A hyperlink in the bottom-right corner of the page connects to Elsevier's official data policy.

2.2 Data Policy Types

Except for Elsevier, Springer Nature, Wiley, SAGE Publishing, and Taylor & Francis have all developed tiered data policies with varying degrees of sharing requirements for their journals to choose from (see -). Although the criteria and number of policy types differ among publishers, all represent a transition from encouragement to mandate, with increasingly stringent requirements for author data sharing. Using Springer Nature as an example (see), its data policy types include four categories. "Data storage in repositories" and "allowing cited datasets to be listed in references" are mandatory across all four types. Beyond these two requirements, Type 1 imposes no other obligations; Type 2 offers optional dataset identifiers and data availability statements but does not mandate them; Type 3 requires authors to provide dataset identifiers and data availability statements; and Type 4 imposes the most stringent requirements, mandating data citation and data peer review.

Wiley's "encourages data sharing," SAGE Publishing's "supports and encourages data sharing," Taylor & Francis's "basic policy," and Springer Nature's "Type 1" are relatively lenient, lacking substantive requirements and relying primarily on encouragement and author voluntariness, yet they publicize the benefits and feasible approaches to data sharing. The most stringent policies, such as Wiley's "mandates data sharing + data peer review" and Springer Nature's "Type 4," similarly mandate data storage in repositories, data citation, verification of data links, data availability statements, and data peer review. Building upon these, Taylor & Francis's "open data + FAIR principles" further requires datasets to follow CC0, CCBY, or equivalent licenses. By contrast, SAGE Publishing's most stringent level, "requires data sharing," is relatively lenient, mandating only data repository deposition, data availability statements, and data citation rules, while data peer review is only optional ("may be asked").

2.3 Journal Adoption

Elsevier's portfolio includes 2,761 journals (2,532 after accounting for discontinued or merged titles). Since Elsevier does not tier its data policies, its jour-

nals share similar brief descriptions of “research data” (except for Cell Press journals), with approximately 1,723 journals (68%) explicitly encouraging data sharing (see [Figure 1: see original paper]). Cell Press journals impose more detailed requirements for paper-linked data. For instance, *The American Journal of Human Genetics* requires nucleic acid and protein sequences, single nucleotide polymorphisms, copy number variations, microarray data, and macromolecular structures determined by X-ray crystallography (including structure factors) to be stored in appropriate public repositories and made unconditionally available upon publication. For genotype data, while not mandating public repository deposition, the journal encourages authors to do so and requires them to provide genotype data as supplementary materials or on websites accessible to readers. Before submitting revised manuscripts, authors must submit newly described single nucleotide polymorphism data to appropriate repositories and deposit all copy number variation data from control samples in the Database of Genomic Variants Archive (DGVA) or the Database of Genomic Structural Variation (dbVAR). The journal also encourages authors to deposit data for other materials used in the paper in appropriate repositories.

[Figure 2: see original paper]-[Figure 5: see original paper] present statistical results of data policy adoption across journals from each publisher. The figures reveal that, except for SAGE Publishing (47.0%), over 68% of journals from the other publishers have explicitly formulated data policies, predominantly encouragement-based. The proportions of journals adopting encouragement policies are 37.6% for Springer Nature, 75.5% for Wiley, 46.9% for SAGE Publishing, and 59.1% for Taylor & Francis. Springer Nature has a higher proportion of journals (approximately 60.1%) adopting transitional levels (Types 2 and 3) compared to Wiley (14.0%), SAGE Publishing (0.8%), and Taylor & Francis (1.3%), indicating that Springer Nature journals impose higher requirements for data sharing. However, SAGE Publishing has the highest proportion of journals (11.5%) that require mandatory data sharing, while other publishers’ proportions are below 3%, likely reflecting differences in subject distributions among their portfolios.

[Figure 2: see original paper] Data Policy Selection Among 1,668 Springer Nature Journals

Note: “Awaiting confirmation” indicates journals that have not yet clearly specified their data policy type; data source: <https://www.springernature.com/gp/authors/research-data-policy/journal-policies-and-services>

[Figure 3: see original paper] Data Policy Selection Among 1,797 Wiley Journals

Note: “Other” indicates journals that have not clearly specified their data policy type or use their own policies; data source: <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/author-compliance-tool.html>

[Figure 4: see original paper] Data Policy Selection Among 1,174 SAGE Publishing Journals

Note: “Other” indicates journals that have not clearly specified their data policy

type or use their own policies; 232 journals support authors uploading supplementary materials to the journal website for publication alongside articles

[Figure 5: see original paper] Data Policy Selection Among 2,911 Taylor & Francis Journals

Note: “Other” indicates journals that have not clearly specified their data policy type or use their own policies

Further statistics were compiled for journals adopting the most stringent data policies (requiring data peer review) (see). SAGE Publishing’s most stringent policy does not require data peer review, so it is not included in this table. The results show that Springer Nature, Wiley, and Taylor & Francis have 9, 2, and 6 journals respectively with the most stringent policies. Except for *Scientific Data*, *Geoscience Data Journal*, and *Big Earth Data* (which are data journals), the others are traditional journals, yet all 17 journals are open access or partially open access, suggesting that traditional non-open access journals face challenges in human, material, and financial resources for data auditing, tracking, and preservation. Notably, 5 of Taylor & Francis’s 6 journals are in geoscience (see), indicating that geoscience, particularly geo-information science, is leading the way in data sharing.

Journals with the Most Stringent Data Policies by Publisher

3. Exploration of Data Policy Formulation and Implementation for Chinese Sci-tech Journals

The experiences of the five international publishers demonstrate that journal data policies play a crucial role in promoting the integrative development of data publication and journals. Therefore, to advance this integration in China, the role of journals must be fully leveraged. Chinese sci-tech journals can learn from international publishers’ experiences and promote data policy formulation and implementation through three aspects: operational mechanisms, policy content, and pilot demonstrations.

3.1 Data Policy Formulation

Chinese sci-tech journals are currently at the stage of establishing whether data policies exist. At this stage, the primary focus should be on promoting awareness and encouraging authors to engage in data publication. Drawing from international experiences, the first step involves disseminating knowledge about data publication and its benefits, followed by tiered requirements for data storage, data citation, and data availability statements, allowing journals to select appropriate levels based on their capabilities. Regarding specific provisions, Springer Nature’s data policy framework is the most comprehensive (see) and can serve as a reference, though the number of tiers and options may be adjusted. For instance, the most lenient policy type might not mandate data storage in repositories. However, we recommend that even the most lenient policy should

require data availability statements, as these merely describe whether paper-linked data are accessible, where, and under what conditions—this does not increase authors’ psychological burden or workload but can enhance awareness of data publication among both authors and readers.

Types and Examples of Data Availability Statements

3.2 Implementation Model Exploration

Chinese sci-tech journals are characterized by being small, dispersed, and weak. At a time when most domestic authors have not yet developed data sharing awareness, relying on individual journals to formulate data policies would be non-standardized, unsystematic, and difficult to advance. Therefore, effective implementation of journal data policies requires not only improved infrastructure, article structures, operational models, and evaluation methods, but also strong coordination among data repositories, sci-tech journals, and other stakeholders. As illustrated in [Figure 6: see original paper], the implementation model comprises four interconnected and coordinated systems: the “Data Publication and Sci-tech Journals Integrative Development Action Group,” organizing and implementation units, data repositories, and sci-tech journals (including data journals).

[Figure 6: see original paper] Implementation Model for Journal Data Policies

3.2.1 Macro-level Regulation by the Action Group The “Data Publication and Sci-tech Journals Integrative Development Action Group” is a key component of the implementation model, with major members including the Publicity Department of the CPC Central Committee, the China Association for Science and Technology, leaders in charge of relevant work, experienced society executives, national experts in sci-tech journal publishing, and data experts. The Action Group is responsible for leading, organizing, and coordinating the national integrative development of data publication and sci-tech journals; researching and formulating deployment plans to advance this integration; promoting coordination and cooperation across departments to accomplish established goals and tasks; supervising and inspecting policy implementation and task completion by relevant units in various professional fields; coordinating the resolution of major issues in implementation; and determining annual work priorities and phased plans while regularly summarizing and reporting progress.

3.2.2 Bridging Role of Organizing and Implementation Units Organizing and implementation units serve as the bridge connecting top-level and ground-level operations, executing the Action Group’s strategic plans. Major organizing units include influential societies with publishing experience such as the Chinese Geographical Society, Chinese Medical Association, and China University Sci-tech Journals Association, as well as host institutions of the Excellence Action Plan for China’s Sci-tech Journals cluster pilot projects. Their primary responsibilities involve translating the Action Group’s strategic vision

into concrete, executable implementation plans; organizing and constructing cooperation mechanisms between sci-tech journals and data repositories; facilitating negotiations between experts from both sides to formulate data policies; providing training, guidance, and supervision to data repositories and journals; evaluating expected outcomes; and regularly reporting to the Action Group.

3.2.3 Division of Labor Between Data Repositories and Journals

Data repositories' most fundamental function is providing hardware and software infrastructure for data storage to ensure secure, permanent preservation. Therefore, infrastructure must be strengthened, standardized data service systems constructed, and comprehensive, analytical, systematic information storage environments established to achieve interconnectivity between data storage and sci-tech journals (including data journals). Repositories must clarify data copyright information, provide unique persistent identifiers for each dataset, establish data review mechanisms to ensure data quality, develop data citation standards to protect data authors' rights and interests, and establish data impact evaluation 指标体系 to enhance data sharing and influence.

Sci-tech journals are the specific implementers of data policies. They should select appropriate policy types based on their disciplinary characteristics from those formulated by organizing units, and adapt them to local conditions. Journals must upgrade their submission systems and article templates to incorporate data policy-related content such as data availability statements and dataset citation formats; leverage the influence of editorial boards, especially editors-in-chief, to publicize journal data policies; investigate authors' opinions on data publication and problems encountered during the process, addressing issues promptly to gradually enhance authors' willingness to publish data; and track changes in metrics such as submission volume, citation counts, numbers of paper-linked data publications, and journal social impact after policy implementation, providing timely feedback to organizing units for adjustments and gradually improving policy adaptation in China.

4. Conclusion

A strong sci-tech nation cannot be built without support from world-class sci-tech journals. In August 2019, four ministries and commissions including the China Association for Science and Technology jointly issued the "Opinions on Deepening Reform and Cultivating World-class Sci-tech Journals," explicitly proposing to explore new publishing models such as data publication. The integrative development of data publication and sci-tech journals is essential for promoting scientific data sharing in China and for building world-class journals. Data policies are important instruments for journals to advance data publication. China's publishing community should seize the opportunity presented by national emphasis on data publication and journal development, learn from the practical experiences of international publishers such as Elsevier and Springer Nature, formulate data policies and operational models suitable for Chinese

sci-tech journals, and earnestly promote the publication and sharing of data associated with journal articles to achieve a win-win outcome.

Acknowledgments: The author thanks Liu Chuang and He Shujin from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, and Gu Liping and Zeng Yan from the National Science Library, Chinese Academy of Sciences, for their valuable suggestions on improving this paper.

References

- [1] Li Hongxing, Wu Lizong, Nan Zhuotong, et al. Research on the model construction of scientific data joint publication [J]. *Remote Sensing Technology and Application*, 2016, 31(4): 801-808.
- [2] Wang Dandan. Data papers: Research on independent publication and sharing models for datasets [J]. *Information and Documentation Services*, 2015(5): 95-98.
- [3] Ma Jianling, Cao Yuezhen, Wang Sili, et al. Research on integrated publication of academic papers and scientific datasets [J]. *Information and Documentation Services*, 2015(2): 82-86.
- [4] Wu Jinming, Qian Qing, Cui Jiawei, et al. Research on the linkage methods between data repositories and journals in data publication [J]. *Journal of Medical Informatics*, 2019, 40(9): 70-76.
- [5] Liu Ziheng, Tu Zhifang. Review of data publication and its quality control [J]. *Library Tribune*, 2020, 40(10): 99-107.
- [6] Global Change Research Data Publishing & Repository [EB/OL]. [2021-03-28]. <http://www.geodoi.ac.cn>.
- [7] Guo Huadong. Foreword to *China Scientific Data* [EB/OL]. [2021-03-28]. <http://www.csdata.org/paperView?id=55>.
- [8] Geoscience Data Publishing Center. Submission guidelines [EB/OL]. [2021-03-28]. <http://geodb.cgs.gov.cn/>.
- [9] *Journal of Agricultural Big Data* [EB/OL]. [2021-03-28]. <http://agbigdata.aiijournal.com/CN/2096-6369/home.shtml>.
- [10] *Big Earth Data* [EB/OL]. [2021-03-28]. <https://www.tandfonline.com/toc/tbed20/current>.
- [11] Prospects for data management and publication in professional academic journals—On the launch of the “Data Papers” column [J]. *Library Journal*, 2018, 37(8): 83.
- [12] *Journal of Remote Sensing*. Guidelines for data paper submission [EB/OL]. [2021-03-28]. http://www.jors.cn/jrs/ch/reader/view_{news}.aspx?id=

- [13] Ball A, Duke M. How to cite datasets and link to publications [EB/OL]. [2021-03-28]. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>.
- [14] China Association for Science and Technology. *Blue Book of China's Sci-tech Journals Development (2019)* [M]. Beijing: Science Press, 2019: 244.
- [15] He Xiaoyan, Wang Jing, Zhang Fangying, et al. Review of Chinese sci-tech journals development in 2020 [J]. *Science-Technology & Publication*, 2021(3): 67-77.
- [16] American Association for the Advancement of Science. *Science*: editorial policies [EB/OL]. [2021-03-28]. <http://www.sciencemag.org/authors/science-editorial-policies>.
- [17] Elsevier. Sharing research data [EB/OL]. [2021-03-28]. <https://www.elsevier.com/authors/tools-and-resources/research-data>.
- [18] Springer Nature. Research data policies [EB/OL]. [2021-03-28]. <https://www.springernature.com/gp/authors/research-data-policy>.
- [19] Fu Tianzhen, Chen Miaozen. Analysis and suggestions on data publication policies of Chinese academic journals [J]. *China Publishing Journal*, 2014(23): 31-34.
- [20] Liu Ying, Wang Xuan. Analysis of data publication policies of international academic journals in the medical field [J]. *Chinese Journal of Scientific and Technical Periodicals*, 2017, 28(8): 685-689.
- [21] Chen Quanping. Research on data policies of academic journals and related studies [J]. *Library and Information Service*, 2015(5): 9-15.
- [22] Kong Lihua, Xi Yan, Zhang Xiaolin. Trends, mechanisms, and challenges in data publication [J]. *Bulletin of National Natural Science Foundation of China*, 2019(3): 237-245.
- [23] *Examples and Counterexamples* [EB/OL]. [2021-03-28]. <https://www.journals.elsevier.com/examples-and-counterexamples/>.
- [24] Wiley. Wiley's Data Sharing Policies [EB/OL]. [2021-03-28]. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>.
- [25] SAGE Publishing. Research Data Sharing Policies [EB/OL]. [2021-03-28]. <https://uk.sagepub.com/en-gb/asi/research-data-sharing-policies>.
- [26] Taylor & Francis. Understanding our data sharing policies [EB/OL]. [2021-03-28]. <https://authorservices.taylorandfrancis.com/data-sharing-policies/>.
- [27] Elsevier. Journals [EB/OL]. [2021-03-28]. <https://www.elsevier.com/catalog?producttype=journals>.
- [28] *The American Journal of Human Genetics*. Information for authors [EB/OL]. [2021-07-01]. <https://www.cell.com/ajhg/authors>.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.