

A Survey and Analysis of Data Policies in Foreign Academic Journals

Authors: Wu Rong, Gu Liping, Liu Jingjing

Date: 2016-06-03T00:00:00+00:00

Abstract

[Purpose/Significance] Academic journals play an important role in research data sharing. By exploring the implementation details of data policies of foreign academic journals, this study provides effective reference for implementing data policies of Chinese academic journals. [Method/Process] Based on pragmatist methodology, analysis is conducted through literature review, and policy elements are screened and refined using rational alternative methods to analyze data policies of academic journals. [Results/Conclusion] The findings indicate that academic journals should formulate data submission policies (including types, formats, and methods), data review policies (including locations where data can be reviewed and verified, and their supporting role for academic papers), and data rights policies (including attribution rules, usage licenses, authors' scientific ethics statements, etc.). It is recommended that in future practice, implementation details suitable for data policies of Chinese academic journals be gradually explored.

Full Text

Preamble

Volume 59, Issue 7, April 2015

Investigation and Analysis of Data Policies in Foreign Academic Journals

Wu Rong^{1,2}, Gu Liping¹, Liu Jingjing^{1,2}

¹ National Science Library, Chinese Academy of Sciences, Beijing 100190

² University of Chinese Academy of Sciences, Beijing 100049

Abstract

[Purpose/Significance] Academic journals play a crucial role in research data sharing. By examining the implementation details of data policies in foreign aca-

demic journals, this study provides effective references for implementing data policies in Chinese academic journals. **[Method/Process]** Based on pragmatist methodology, we conducted a literature review analysis and used a reasonable competing alternatives approach to screen and extract policy elements for analyzing academic journal data policies. **[Result/Conclusion]** The findings indicate that academic journals should formulate data submission policies (including data types, formats, and submission methods), data review policies (including data reviewability, verifiable storage locations, and support for academic papers), and data rights policies (including attribution rules, usage licenses, and authors' scientific ethics statements). We recommend that future practice gradually explore implementation details suitable for Chinese academic journal data policies.

Keywords: Academic Journals; Data Policy; PLoS; Data Sharing; Desktop Research

Classification Number: G231

DOI: 10.13266/j.issn.0252-3116.2015.07.013

1. The Trend of Promoting Open Data Sharing in Academic Communication

Scientific progress depends on scientific ideas and evidence. The Royal Society's report "Science as an Open Enterprise" proposes that a complete academic paper should include a full description of experiments, result data, uncertainty evaluation, and metadata to ensure data can be verified and reused. However, over the past 30 years, the publication of raw data and scientific results has become disconnected [1]. Due to length limitations, current academic papers typically only present academic ideas and discoveries, while scientific data containing scientific theories often remains buried in scattered laboratory notebooks [2]. To enable reproducible verification and further research, it is necessary to effectively trace the research data supporting scientific findings. Unlike academic papers, the generation, preservation, management, and dissemination of research data involve numerous complex factors that must be considered from the perspectives of various stakeholders, requiring an overall policy framework before examining each aspect individually [3].

Research data storage and management represent an important trend in scientific development, yet many urgent issues remain in practice. Among foreign scholars, S. Kutay [4] discussed issues such as digitization, copyright, and bulk downloading that research institutions must address when developing research data resources. C. M. Perry [5] surveyed Canadian researchers on handling research data from publicly funded projects and found that researchers have begun to recognize ethical issues in research activities. P. Arzberger et al. [6] concluded that researchers' data-sharing behaviors are influenced by technical, organizational, financial, legal, and policy factors. Domestically, Tang Yi et al. [7] reviewed policies and regulations on scientific data, including journal policies for scientific data sharing.

Based on these factors affecting research data dissemination, scholars have summarized various methods for disseminating research data [8], as shown in Table 1.

Table 1. Research Data Dissemination Methods

Data Publication: Under copyright agreements, data is published on the internet or other media with citation and reuse provisions. Currently, many international data repositories support research data publication.

Data Materials: Provided as supporting materials for academic papers, either as paper attachments or stored in data repositories.

Data Journals: Focus on describing experimental and observational datasets, published as data descriptors.

As a medium for research data dissemination, many foreign academic publishers and journals, including *Science* [9], *PNAS* [10], *Nature* [11], and *BMC* [12], have explicitly proposed policies for research data storage and dissemination. PLoS, as one of the few journals practicing open sharing and among the earliest to propose research data policies, has had broad influence and relatively standardized implementation. Since March 1, 2014, PLoS has implemented a new research data open policy requiring authors to submit a data availability statement based on PLoS policies when submitting papers, sharing research data and related metadata and methods [13]. Currently, Chinese academic journals are at an early stage in research data storage and management. Studying the research data policies of renowned foreign journals such as PLoS offers important lessons for Chinese academic journals. Considering the current state of Chinese academic journals, this paper proposes strategic recommendations for their data policies.

2. Research Framework for Academic Journal Data Policies

Academic journals are continuously deepening and refining their research data management measures, yet several issues remain in implementation. For instance, what specifications must authors follow when submitting data? How should journals process submitted data? What stakeholders' rights and interests must be considered in this process? These issues translate into formal research questions: When submitting data, must authors determine whether journals require mandatory data submission and what data types are required? What are the attribution requirements for data contributors? Are there differences in data openness scope and timing? What are the regulations on data citation, storage, and dissemination? These questions are crucial for research data management and require policy guidance.

This study's methodology references the process from "policy content collection" to "policy recommendation proposal" mentioned in *Introduction to Foresight—Promoting Policy and Shaping the Future* [14]. Using a policy observation approach, we established policy observation priorities and designed a research

framework dividing academic journal data policies into three research questions, each including several observational issues, as shown in Table 2 .

Table 2. Analysis Framework for Academic Journal Data Policies

Research Question 1: What specifications must be followed when submitting data? - Is data submission mandatory when submitting papers? - What data should be submitted? What data can be submitted but is not mandatory? - Are there requirements for format and citation methods? - What are the submission methods and priorities? If other methods are allowed, what are the specific requirements?

Research Question 2: How do journals review and preserve submitted data? - Is data provided only to reviewers or must it be publicly available? - Do journals review submitted data? How? - Are there explicit regulations for long-term data preservation? - If storing in public databases, are there requirements? Are designated databases specified? - Do journals provide data services to authors?

Research Question 3: What rights issues must stakeholders consider? - How are data copyrights regulated? - How are data usage licenses regulated? - What requirements exist regarding scientific ethics, privacy protection, and third-party rights?

Based on this framework and referencing PLoS' s Editorial and Publishing Policies [13], as well as *Science*' s Data and Materials Availability [9], *PNAS*' s Editorial Policies [10], *Nature*' s Data Policies [11], and BMC' s position statement on open data [12], we conducted a detailed analysis of each policy observation.

3. Data Submission Specifications

3.1 Mandatory Data Submission

PLoS data policies explicitly state that, with very few exceptions, authors must provide unrestricted open access to all data related to research results. When submitting papers, authors must submit a data availability statement based on PLoS policies, indicating that data has been publicly stored and listing the repository name and data DOI. If authors cannot obtain a DOI before acceptance, they must provide it upon acceptance; otherwise, the paper will not be published [13]. If authors refuse to share data and related metadata and methods, their papers will be rejected. PLoS reserves the right to issue corrections or, in extreme cases, retract publications if data access remains restricted after publication. Papers will not be considered if authors refuse data sharing for personal gain or when research conclusions depend entirely on commercially interested or copyright-protected proprietary data analysis [13]. *Science* [9], *Nature* [11], and other academic journals currently require authors to store data forming paper conclusions when submitting papers and provide complete open sharing after publication. *PNAS* [10] requires authors to store data when submitting papers and disclose limitations on material and information availability.

Chinese academic journals should make data submission a mandatory condition for publication during open data implementation. This not only facilitates standardized data management and avoids redundant investment but also lays the foundation for verifying, reproducing, and further researching paper conclusions, enhancing research reliability and diversity.

3.2 Types of Submitted Data

PLoS data policy defines mandatory submitted data as the “minimal dataset,” which includes datasets and related metadata and methods used to form paper conclusions, plus any additional data needed to fully reproduce reported results. Core descriptive data, methods, and research results should be included in the paper body regardless of data storage. All data and related metadata should be stored in an appropriate public database unless already provided as part of the paper [15]. The U.S. National Institutes of Health (NIH) defines such data as “final data,” meaning materials generally considered necessary by the academic community to verify research results [16]. Journals require authors to submit data that directly forms paper conclusions but do not mandate raw, unprocessed data, though authors are encouraged to store it in suitable repositories.

Chinese academic journals should require authors to submit data types referencing foreign journals’ “minimal dataset.” This ensures data reusability while preventing data overuse and maximizing author rights protection.

3.3 Data Format Requirements

Due to disciplinary differences in data formats, PLoS data policies recommend authors follow domain-specific requirements and choose domain-appropriate data repositories, submitting data in repository-required formats. For data uploaded as paper attachments, authors should select file formats enabling effective data extraction to maximize accessibility and reusability. For example, Excel format is preferable to PDF for tabular data; if main results are images, authors should submit data tables generating those images [15]. BMC data policies similarly state that to maximize data reuse potential and improve research efficiency, shared data formats should adopt widely recognized domain standards. For instance, microarray data should follow MIAME format, molecular interaction data should follow MIMIx guidelines [10-11], and mass spectrometry data should be provided in mzML format [17].

Given significant disciplinary differences, journals cannot uniformly require data formats, and authors must comply with repository requirements when selecting storage locations. Uniform format requirements are impractical in policy implementation; editors and reviewers should require specific articles to use domain-appropriate formats based on circumstances. Table 3 lists common data formats in remote sensing [18], medicine [19], biology [20], and geography [21].

Table 3. Common Data Formats by Discipline

Remote Sensing: HDF5, GeoTIFF, FAST, NDF, CEOS, DIMAP...

Medicine: SCP, HL7 aECG, DICOM, FDAXML...

Biology: FASTA, FASTQ, GenBank, EMBL, GFF...

Geography: DGN, HDF, NITF, TIFF, SDC, DLG, ETAK, GIRAS, IGDS, IGES
...

3.4 Data Submission Methods

PLoS data policies require authors to submit a data availability statement when submitting papers, clearly indicating data has been stored in public repositories and providing data DOIs. Repositories like Dryad, a PLoS integrated partner, generate review URLs for peer review, enabling private data access before publication. If manuscripts are accepted, papers and datasets are published simultaneously [15]. However, not all repositories provide peer-review access mechanisms [10]; if unavailable, authors can provide data through other means, such as email with compressed files [13].

Unlike foreign publishers that mostly recommend large public repositories, Chinese academic journals (except geography journals with designated storage [22-24]) should recommend storing data in open-access institutional repositories with accurate links. PLoS' s proposed submission method is a general but not absolute solution requiring case-specific adjustments. Authors may also store data in institutional repositories and submit storage locations to editorial offices as open data pathways.

4. Data Review and Storage Methods

4.1 Data Openness Timing and Scope

PLoS requires submitted data to be fully open without restriction after publication; data remains closed before paper publication. Many repositories allow private access for peer-review purposes [13]. For repositories not supporting review, authors may submit specific data types to editors through alternative methods. Therefore, journals should list permitted data formats by discipline for author reference. During review, only editors and reviewers can access paper-related data; before publication, data is only available to peer reviewers [15].

Chinese academic journals should, when opening data to protect authors' rights and prevent pre-publication overuse, limit data access to editors and reviewers during the review period and open data for sharing only after paper publication.

4.2 Data Review

To ensure data quality and usability, PLoS reviews submitted data. Using Dryad as an example, after authors submit data, a review URL is generated and uploaded to the submission system as a password for private data access before publication. Manuscripts then proceed rapidly through peer review. If accepted, papers and datasets are published simultaneously [15]. However, some foreign

journals lack review mechanisms, delegating data review entirely to repositories, though this is rare in observed policies.

Chinese academic journals should require submitted data to support paper conclusions, undergo multi-party verification, and contain no fabrication; otherwise, paper acceptance will be directly affected.

4.3 Long-term Data Preservation

PLoS data policies do not explicitly require long-term preservation, 原则上 respecting author institutions' and funders' requirements. Many research institutions and funders have specific requirements, such as the UK Research Councils (RCUK) requiring data preservation for over 10 years after project completion with minimal restrictions [22]. *Nature* requires authors to preserve local datasets for at least 5 years.

Foreign publishers' approaches to long-term preservation fall into three categories: (1) Journals hold authors responsible, with preservation depending entirely on authors and stakeholders, while repositories guarantee storage for certain years; (2) Journals delegate preservation to repositories, providing repository lists with corresponding preservation terms without direct requirements; (3) 极少数 journals have designated cooperative repositories, such as NPG data stored in *Scientific Data* for long-term preservation meeting journal-specific requirements [9].

Chinese academic journals have unique advantages in long-term preservation. Using Chinese Academy of Sciences journals as an example, each institute has an independent institutional repository requiring authors to store data in institute-based data repositories when submitting to affiliated journals. Under research funders' and institutions' preservation policies, authors should provide retrievable and usable institutional repository locations.

4.4 Data Repositories

PLoS accepts domain-specific or thematic proprietary databases accepting particular data types, as well as general databases accepting multiple data types. Recognizing data diversity and different repository functions, PLoS requires authors to follow domain-specific storage standards and choose appropriate databases. PLoS encourages authors to select trustworthy repositories meeting acceptance standards, such as those approved by the Centre for Research Libraries or Data Seal of Approval [13]. For example, microarray data in ArrayExpress or GEO, gene sequence data in GenBank, EMBL, or DDBJ, and ecology data in Dryad. For journals without specialized data, authors should store data in general databases like Dryad or Supplementary Material [23].

Large international databases are more likely to persist long-term than small local databases. Under Chinese circumstances, institutional repositories should

primarily store research data, with journals recommending disciplinary repositories when institutional storage is unavailable.

4.5 Data Services

As PLoS' s first integrated partner, Dryad supports all PLoS journal data, making research data discoverable, freely reusable, and citable. Dryad connects with all PLoS journals, integrating data storage with paper submission. PLoS data services also include recommending repository lists, data quality review, and providing data DOIs [15]. BMC collaborates with LabArchives to offer authors 100MB of free storage after manuscript submission [27]. *Science* [9], *Nature* [11], and BMC [12] provide optional repository lists by discipline.

Chinese academic journals should provide data services including repository recommendations and data review, and may offer ORCID and institutional repository links for author and data identification, plus data backup, format conversion, and unified description services based on actual conditions.

5. Data Rights Management

5.1 Data Copyright Ownership

PLoS data policies note that copyright licenses in data repositories remain unclear [13]. Discrepancies may exist between paper authors and data authors—for example, paper authors may be individuals or groups, while data may come from institutions rather than individuals, or author order may differ between papers and data. Conflicts in attribution and ordering require clear regulations, though currently difficult to specify. Most attribution rules should follow research institution or funder regulations, with journals only generally addressing attribution. Currently, when Chinese researchers submit to foreign journals, China' s lack of relevant regulations while foreign publishers have storage and sharing requirements results in data loss abroad, causing Chinese institutions and personnel to lose data initiative.

Chinese academic journals should determine copyright ownership based on research funder project requirements, institutional work-for-hire requirements, and scientific ethics constraints. Authors should clarify attribution when submitting papers and data and allow journals to disseminate and utilize data under usage license conditions.

5.2 Data Usage Licenses

PLoS data policies state that if authors choose repositories with specified storage license policies, these should not be stricter than CC-BY. Researchers using data must comply with domain-specific ethical norms [13]. BMC states that unless otherwise noted, data licenses are under CC-BY [12]. The Royal Society' s *Royal Society Open Science* [26] also recommends CC-BY. Due to journals' commercial

nature, they rarely have NC (non-commercial) attributes and mostly adopt CC-BY licenses. Data license regulations vary by domain and journal and should be studied holistically by science and technology information policy research teams for repositories, funders, journals, institutional repositories, and researchers.

To respect research funders' and institutions' data management norms and since academic journals are not actual research management units, journals should require authors to provide data rights statements, with templates provided by publishers. Academic journal data usage licenses include four forms: CC0, CC-BY, CC-BY-NC, and CC-BY-NC-ND. Unless otherwise specified, CC-BY should be the default license, protecting author attribution rights while allowing content reuse, data mining, and commercial use by readers, the public, researchers, journals, and third parties.

5.3 Privacy Protection and Rights Requirements

PLoS data policy exceptions for open storage and sharing include data involving law, privacy, and third-party data. For such data, PLoS proposes two alternative solutions requiring case-by-case evaluation [13].

5.3.1 Privacy-Involving Research Data When research data involves privacy, sharing scope can be limited to researchers with study needs. Data availability statements must explicitly state “data access is based on reasonable request” and indicate which committee (e.g., data access or ethics committee) should receive requests. Reasons for limiting public storage must be detailed. For human subjects research, PLoS recommends researchers handle data according to existing guidelines and local laws to ensure participant privacy. Privacy protection steps may include identity authentication, restricting data to specific databases, or clarifying privacy-related license agreements. As part of the statement, authors should indicate privacy protection methods. If license agreements apply, authors should note that permission from other researchers is required. Importantly, authors cannot be solely responsible for research data access.

5.3.2 Third-Party Data If the primary dataset does not come from submitting authors but uses third-party data, interested researchers must obtain it from designated original sources. In this case, data availability statements must use complete citations to indicate data sources. If datasets cannot be provided, they should display “data from [named source].” Reasons for limiting public storage must be detailed.

Addressing these three research questions and referencing *Introduction to Personas—Online User Behavior and Differentiated Services* [28] Chapter 2, we used pragmatist methodology's four principles to determine specific content selection and used reasonable competing alternatives to identify optimal solutions, extracting policy elements to form data submission specifications, data review policies, and data rights policies, as shown in Table 4 .

Table 4. Core Elements of Academic Journal Data Policies

Data Submission Mandate: Data must be submitted with papers to ensure authenticity and usability.

Data Types: Submit processed data related to paper conclusions, not all raw data.

Format Requirements: Follow domain-specific requirements (e.g., ASCII, BMP, HDF in remote sensing; SCP, DICOM in medicine).

Submission Method: Store in open-access institutional repositories with accurate links.

Openness Timing & Scope: Open to reviewers during review; fully open after publication.

Data Review: Data must support conclusion validation; otherwise affects acceptance.

Long-term Preservation: Provide retrievable institutional repository URLs under funder/institution management.

Repositories: Primarily institutional repositories; may use journal-recommended disciplinary repositories.

Data Services: Recommend repository lists, conduct reviews, provide ORCID/institutional links, backup, format conversion, unified description.

Copyright Ownership: Determine based on funder requirements, institutional work-for-hire rules, ethics constraints; authors must clarify attribution.

Usage Licenses: Default to CC-BY unless otherwise specified, protecting attribution while allowing reuse.

Privacy & Rights: Data involving law, privacy, or third-party data requires case-by-case evaluation.

6. Conclusion

Numerous challenges exist in scientific data open sharing, and academic journal data policies represent an attempt and progress toward solutions [29-30]. Good research data policies can reduce sharing costs in research data management [31]. Many foreign academic journals have begun participating in research data management, integrating paper publication with open data sharing. Chinese academic journals have not yet achieved scale in research data management and must actively engage in management and services. While learning from foreign good practices and considering Chinese academic journals' actual conditions, they should organically combine paper publication with open data sharing to seize opportunities and gain initiative in the new academic environment.

References

- [1] Science as an open enterprise [EB/OL]. [2015-01-21]. https://royalsociety.org/~media/royal_society_content/06-20-saoe.pdf.
- [2] Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery [M]. Translated by Pan Jiaofeng, Zhang Xiaolin, et al. Beijing:

Science Press, 2012.

[3] Gu Liping. Basic Framework for Scientific Data Rights Analysis [J]. Library and Information Science, 2014(1): 34-51.

[4] Kutay S. Advancing digital repository services for faculty primary research assets: An exploratory study [J]. The Journal of Academic Librarianship, 2014, 40(6): 642-649.

[5] Perry CM. Archiving of publicly funded research data: A survey of Canadian researchers [J]. Government Information Quarterly, 2008, 25(1): 133-148.

[6] Arzberger P, Schroeder P, Beaulieu A, et al. Promoting access to public research data for scientific, economic, and social development [J]. Data Science Journal, 2004, 3: 135-152.

[7] Tang Yi, Zhang Xiaomeng, Zheng Ran. International scientific data sharing policy and regulation system: Linked Science institutional foundation [J]. Library and Information Knowledge, 2013(3): 67-73.

[8] Liu Fenghong, Cui Jinzhong, Han Fangqiao, et al. Data papers: Discussion on emerging academic publication types in the big data era [J]. Chinese Journal of Scientific and Technical Periodicals, 2014, 25(12): 1451-1456.

[9] Science. Data and materials availability [EB/OL]. [2015-01-21]. http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail.

[10] PNAS. Editorial policies [EB/OL]. [2015-01-21]. <http://www.pnas.org/site/authors/journal.xhtml>.

[11] Nature. Data policies [EB/OL]. [2015-01-21]. <http://www.nature.com/scientificdata/for-authors/data-deposition-policies/>.

[12] BioMedCentral. BioMedCentral's position statement on open data [EB/OL]. [2015-01-21]. <http://blogs.biomedcentral.com/bmcblog/files/open-data-statement-draft.pdf>.

[13] PLoS. PLoS editorial and publishing policies [EB/OL]. [2015-01-21]. <http://www.plosone.org/static/policies.action>.

[14] Gu Liping. Introduction to Foresight—Promoting Policy and Shaping the Future [M]. Taipei: Designer Press, 2013.

[15] Lin J. Make data sharing easy: PLOS launches its data repository integration partner program [EB/OL]. [2015-01-21]. <http://blogs.plos.org/tech/make-data-sharing-easy-plos-launches-its-data-repository-integration-partner-program/>.

[16] National Institutes of Health. NIH data sharing policy and implementation guidance [EB/OL]. [2015-01-21]. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.h

[17] BioMedCentral. Availability of supporting data [EB/OL]. [2015-01-21]. <http://www.biomedcentral.com/about/supporting-data>.

[18] Guo Jing. Foreign remote sensing data format standards and implications [J]. Aerospace Standardization, 2011(4): 29-31.

[19] Qiu Sihai, Zeng Yonghua. Analysis of standard data formats in ECG data management systems [J]. Progress in Biomedical Engineering, 2010(2): 63-67.

[20] Common biological databases and data formats [EB/OL]. [2015-01-21]. <http://www.docin.com/p-659623271.html>.

[21] ArcDIS Resources. About geographic data formats [EB/OL]. [2015-01-21]. <http://resources.arcgis.com/zh-cn/help/main/10.1/index.html#/018m0000000020000000>.

[22] Zhu Yunqiang, Sun Jiulin, Liao Shunbao, et al. Research and practice

- on earth system science data sharing [J]. Journal of Geo-Information Science, 2010, 12(1): 1-8.
- [23] Liu Rong, Zhang Na. Sharing reflects scientific data value—Interview with academician Sun Jiulin [J]. Science and Technology Innovation and Brands, 2011(7): 10-13.
- [24] Fu Xiaofeng, Li Jun, Li Jianhui. International scientific data development and sharing [J]. China Basic Science, 2007(2): 30-35.
- [25] National Science and Technology Infrastructure Center. National Science and Technology Infrastructure Development Report (2011-2012) [M]. Beijing: Scientific and Technical Documentation Press, 2013: 248.
- [26] The Royal Society. Royal society open science [EB/OL]. [2015-01-21]. <http://royalsocietypublishing.org/royal-society-open-science>.
- [27] BMC Central. Availability of supporting data [EB/OL]. [2015-01-21]. <http://www.biomedcentral.com/about/supporting-data>.
- [28] Gu Liping. Introduction to Personas—Online User Behavior and Differentiated Services [M]. Beijing: Science and Technology Literature Press, 2013.
- [29] Borgman CL. The conundrum of sharing research data [J]. Journal of the American Society for Information Science and Technology, 2012, 63(6): 1059-1078.
- [30] Qing Xiuling. Challenges in research data sharing [J]. New Technology of Library and Information Service, 2013(5): 1-20.
- [31] Pisani E, Whitworth J, Zaba B, et al. Time for fair trade in research data [J]. The Lancet, 2010, 375(9716): 703-705.

Author Contributions

Wu Rong: Main paper writing, policy detail 梳理;

Gu Liping: Research design, policy element analysis, final paper revision;

Liu Jingjing: Policy document translation.

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

Source: ChinaXiv—Machine translation. Verify with original.