Comparative Study on Practice Models of Data Publishing —A Case Study in the Geoscience Domain (Postprint)

Authors: Han Lu, Ding Yi

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

Abstract

[Purpose / Significance] Scientific data publishing represents the primary scholarly communication mode for data-intensive scientific discovery and holds significant importance for enabling data reuse and scientific verification. [Method / Process] Earth science has undergone tremendous transformation from traditional data sharing models to current data publishing practices. The authors classify data publishing into three modes: data journal publishing, data repository publishing, and integrated data and paper publishing. For each mode, practical methods and key elements are statistically analyzed and compared, with focused analysis on the strengths and weaknesses of the three modes, issues of peer data review, and the importance of hierarchical metadata in geoscience data publishing. [Results / Conclusions] Through this research, the authors propose that repository-based publishing facilitates integration into scientific data management processes and benefits data reuse, yet such publishing modes lack peer review; peer review of data should differ from that of academic papers, emphasizing the reusability of data in the process of participating in scientific research and generating reproducible results; hierarchical metadata description holds significant importance for the preservation and reuse of big geoscience data.

Full Text

A Contrastive Study of Practical Modes of Data Publishing —Take the Field of Earth Science as an Example

Han Lu[1], Ding Yi[2]

[1]Institute of Scientific and Technical Information, Chinese Academy of Geological Sciences, Beijing 100037
[2]Institute of Geology, Chinese Academy of Geological Sciences, Beijing 100037

Abstract

[**Purpose/Significance**] Scientific data publishing is the primary mode of academic communication for data-intensive scientific discovery, which holds great significance for enabling data reuse and scientific verification. [**Method/Process**] Earth science has undergone tremendous transformation from its original data sharing model to the current data publishing paradigm. This paper divides data publishing into three modes: data journal publishing, data repository publishing, and joint publishing of data and papers. For each mode, the authors conducted statistical analysis and comparison of practical methods and key elements, with particular focus on analyzing the advantages and disadvantages of the three modes, issues of peer data review, and the importance of hierarchical metadata in geoscience data publishing. [**Result/Conclusion**] Through this research, the authors propose that repository-based publishing facilitates integration into scientific data management processes and promotes data reuse, though such publishing modes typically lack peer review. Peer review of data should differ from that of academic papers, emphasizing instead the reusability of data in the process of participating in scientific research and generating reproducible results. The hierarchical description of metadata is of great significance for the preservation and reuse of geoscience big data.

Keywords: data publication; data repository; data journal; geoscience data sharing

---

Earth science (hereinafter referred to as "geoscience") is fundamentally a data-driven discipline. However, due to challenges such as difficult data collection, wide spatial coverage, and expensive instrumentation, data sharing, acquisition, and reuse have always been critical issues in geoscience research. In the early 20th century, punch cards were used to record data. During the 1970s, to facilitate geoscience data sharing, numerous international organizations were established, such as the World Data Center (WDC, replaced by the World Data System, WDS, in 2008), the Group on Earth Observations (GEO), and the Data Observation Network for Earth. China joined WDS in 1988 and established nine data centers, most of which were in the field of earth sciences, including seismology, geology, and geophysics data centers. In 2002, the Scientific Data Sharing Project launched the "Earth System Science Data Sharing Network Project," establishing the Geoscience Data Sharing Platform. However, data acquisition from these data centers and sharing platforms was often restricted, sharing activities were inactive, and data silos even formed. Chu Yunqiang et al. studied the sharing mechanisms of scientific big data and proposed that the lack of policies and mechanisms for data sharing was the main obstacle to its development. An important mechanism to motivate researchers to actively share data is bottom-up scientific data publishing.

Data publishing is a new concept jointly proposed by the publishing and data

sharing communities in recent years. Under the trend of open science, funding agencies and international organizations in various countries have introduced policies requiring the management and sharing of research data. For example, the Division of Earth Sciences of the U.S. National Science Foundation (NSF) requires that complete datasets, data products, software, and integrated data must be publicly accessible within two years. Because data centers require users to acknowledge the contributions of data authors when using data, and academic journals require authors to provide data related to scientific conclusions when publishing papers, these two mechanisms have ultimately become the common driving force for data publishing. The core of data publishing is to provide standard citation formats and permanent access addresses for data citation, making scientific data accessible, understandable, evaluable, and usable. Compared with traditional data sharing, data publishing places greater emphasis on the possibility of data reuse and the degree of recognition by researchers. Germany was the first country to study scientific data publishing; the German National Library of Science and Technology applied to become the world' s first DOI registration agency for scientific data and registered DOIs for many data resources from former WDS data centers. These data centers gradually developed data publishing practices, and data publishing has become a new form of sharing. The establishment of a data citation evaluation system objectively reflects the influence of data contributors and improves the previously inactive state of data sharing. Geoscience data publishing in China is still in its infancy, forming a complementary relationship with existing data sharing. The China Geological Survey has already initiated DOI registration for data. Due to the early start of data sharing in the geoscience field, data publishing practices in this domain are representative. Some traditional geoscience journals have long had precedents for publishing data as supplementary materials, such as *Acta Petrologica Sinica*. Through investigating various data publishing modes in the geoscience field, this paper explores the main existing and urgent issues in geoscience data publishing.

Since the concept of data publishing was proposed, some WDS centers have gradually transformed their original data sharing services into data storage management and publishing services, focusing on data discovery, acquisition, reuse, and traceability. This paper primarily examined data centers and repositories registered on the re3data.org platform. Among them, 647 repositories have been registered in the earth sciences field, second only to biology, making it a discipline with numerous data publishing practices.

Using the Journal Citation Reports (JCR), the authors queried journals in geology, geophysics, geochemistry, environmental science, geography, and multidisciplinary earth sciences. The authors also identified data sources in the earth sciences field from Clarivate Analytics' Data Citation Index (DCI) and journal website descriptions. The earth science data journals indexed in JCR mainly include three types: *GeoScience Data Journal* (GDJ), *Scientific Data* (SD), and *Earth System Science Data* (ESSD). In 2014, China launched the Global Change Research Data Publishing & Repository (GCRDPR), a data publishing

platform in the field of earth sciences. In 2017, this publishing system launched the *Journal of Global Change Data*. Some data repositories collaborate with reputable academic journals to increase data influence and achieve interconnection between academic papers and data. Among geoscience data repositories, Pangaea has published a large volume of data. Elsevier' s Science Direct and Scopus have both partnered with Pangaea to achieve integrated publishing of academic papers and data. Another category involves publishing data as supplementary materials with academic papers, but since the main focus of such publishing remains the academic paper itself, this paper does not discuss this mode. In summary, the main data publishing modes in the geoscience field include: (1) Data journals, which transform data into paper format for description and publish them in data journals, usually in conjunction with domain-specific data centers or public data repositories for dataset deposit, with data papers undergoing a peer review process similar to academic papers. (2) Repository-based direct data publishing, which releases data through data repositories, some transformed from original data sharing platforms, with data publishing serving as a form of data management and sharing.

2. Geoscience Data Journals

Data journals differ from traditional academic paper publishing in that they primarily provide and describe datasets themselves, usually without focusing on the scientific innovativeness of conclusions. The main purpose of publishing is to enhance data influence, enabling scientific data to be more thoroughly described, explained, and reused. The authors compared four geoscience data journals (see Table 1), with journal article statistics shown in Figure 1 [Figure 1: see original paper], where SD statistics only include articles under the Earth and Environmental Science themes.

Table 1. Survey of Earth Science Data Journals

	Wiley Online Library (GDJ)	Nature (SD)	GCRDPE	Copernicus Publications (ESSD)
Publisher	Wiley	Nature Publishing Group	*Journal of Global Change Data* Editorial Office	Copernicus Publications

	Wiley Online Library (GDJ)	Nature (SD)	GCRDPE	Copernicus Publications (ESSD)
Journal Positioning	Interdisciplinary journal in earth sciences	Comprehensive journal including Earth and Environmental Science themes	Interdisciplinary journal in earth sciences	Interdisciplinary journal in earth sciences
Peer Review Process	Internal journal peer review	Internal journal peer review	Two-stage open peer review online	Anonymous internal expert review
JCR Impact Factor (2016)	-	4.836	-	8.286
Dataset Access Method	Submit to recommended repository	Submit to recommended data repository	Submit to recommended repository	Data repository provided by journal system
Dataset Citation Standard	DataCite	Custom standard	DataCite	Custom standard
Publication Format	Online and print	Online	Online	Online

2.1 Ensuring Data Accessibility

All journals require data to be submitted to data repositories for long-term preservation and to provide a sustainably citable unique identifier. Currently, most repositories use DOI as the address resolution standard. GDJ, SD, and ESSD do not have their own data storage repositories, so these journals provide lists of recommended repositories where data can be submitted. During paper submission, authors must simultaneously select a recommended repository for data preservation. Most data journals utilize repositories registered

on the re3data.org platform or select national data centers that align with the journal' s content. This publishing model requires cooperation between publishing institutions and data preservation management agencies. GCRDPR differs slightly from the other three journals, as it is itself an online publishing system that integrates metadata, entity data, and data paper publishing. After initial data submission, the publishing platform determines whether to accept the data based on metadata quality and data description. If accepted, the platform notifies the author to write a data paper according to guidelines, assigns a DOI to the dataset, and then initiates the data paper peer review process.

2.2 Writing Standards for Data Papers

Data papers provide enhanced descriptions of dataset contents. Different disciplines have varying dimensions for describing data papers. The authors summarized the content frameworks for dataset description across four geoscience data journals, covering nine aspects: (1) effective dataset access methods, such as providing DOI or URI; (2) dataset coverage, including temporal and spatial coverage; (3) dataset format information, such as data format, encoding method, and encoding language; (4) dataset licensing for legal use; (5) project information providing dataset production details, such as purpose and funding source; (6) provenance information describing dataset production methods, including tools and processing methods used; (7) quality information describing data quality, such as dataset limitations and anomalies; (8) reuse information providing dataset usage descriptions to promote potential reuse; and (9) support and standards for data citation. Table 2 compares the content frameworks of the four journals. ESSD does not explicitly specify descriptive content elements, but during data paper publication, editors advise authors to focus on content that determines the potential value of data papers and provide review guidelines that include content essentially identical to the nine aspects listed in Table 2. The *Journal of Global Change Data* does not have unified specifications for data paper content but instead standardizes these in its metadata.

Figure 1. Statistics of article volumes by journal

Table 2. Comparison of dataset description requirements across four journals

Dataset Description Requirement	GDJ	SD	ESSD	GCRDPR
Dataset accessibility identifier (DOI or URI)				
Dataset coverage				
Dataset format information				
Dataset licensing information				
Data reuse information				
Data quality description				
Data provenance information				
Support and standards for data citation				

2.3 Peer Review and Quality Control

In terms of peer review, SD, GDJ, and GCRDPR adopt peer review processes similar to traditional academic journals, but the review content focuses more on data copyright and quality. GCRDPR provides a specialized review content template covering five aspects: dataset significance, quality, academic relevance, author intellectual property awareness, and data accessibility. ESSD differs slightly from the other three journals by employing a two-stage open peer review process. After authors submit their initial manuscript, it is published online as a discussion paper. Following expert review, professional editors evaluate the manuscript but do not assess its scientific content, focusing instead on whether the data paper meets writing requirements and falls within the journal' s scope, providing technical revision suggestions. Papers remain in this state for eight weeks, during which various peer reviews and discussions can occur, with each paper receiving final evaluation from at least two experts. During the open discussion phase, authors must respond to comments and revise their initial manuscript before a final publication decision is made. For data paper peer review, the publishing community has not yet established unified standards, making this two-stage open review approach more conducive to receiving accurate feedback from actual data users or academic peers.

3. Repository-Based Direct Data Publishing

3.1 Elements of Repository-Based Direct Publishing

Repository-based data publishing emerged from the transformation of original data sharing methods. Currently, no explicit standard specifies required elements across fields. J. E. Kratz' s survey on data publishing found that most researchers focus on whether data has unique identifiers, is openly accessible, is stored in a repository, and has formal rich metadata. Since re3data.org began providing registration services for research data preservation and management repositories worldwide, repository construction has been standardized and repository influence enhanced. Some publishers and journals (such as PeerJ, Springer, and Nature' s *Scientific Data*) use re3data.org as a simple tool to identify suitable data repositories. The authors compared data repositories recommended by data journals in Table 1 and found that common elements for data publishing include the following five aspects: (1) providing unique identifiers for data to ensure 固有 addresses for access; (2) providing data submission editing tools and conducting data quality control, including data submission, data description editing, and metadata editing, with data quality control mostly completed by data administrators to ensure the completeness and accuracy of submitted content; (3) providing formal data usage licensing statements—repositories, lacking commercial copyrights like journals, provide dual protection for data creators and users in an open science environment, with widely adopted licenses including Creative Commons family' s CC0, CC-BY, and CC-BY-SA; (4) providing formal data citation formats, as formal citation is the mechanism for data reuse and location, with T. E. Pronk et al.' s game-theoretic analysis showing that

reducing costs and increasing citations are more incentivizing than policy regulations; and (5) enabling open access to data, as data accessibility is a prerequisite for reuse, with most repositories providing data publishing services being openly accessible, though a few require permissions.

The authors searched re3data.org using these five elements and found that most repositories satisfy open access requirements, with only a few being tiered-access. Regarding data identifiers, 234 repositories currently provide this service, predominantly using DOI, ARK, and URI, with 101 repositories adopting DOI. Repositories not providing this service mostly use external public repository registration services for their data management. For data submission editing and quality control, unlike journals where quality control is conducted by data managers as technical review rather than scientific peer review, two main approaches exist: (1) some repositories with highly specialized data closely related to projects include domain experts for data validation, such as the U.S. National Snow and Ice Data Center; and (2) some repositories support and have obtained certifications, predominantly WDS, DSA (Data Seal of Approval), and CoreTrustSeal, with 48 repositories belonging to WDS. For formal data citation, since data publishing is relatively new, no unified standard exists domestically or internationally, though geoscience data publishing widely references the RORCE11 data citation alliance statement and DataCite' s publishing and citation scheme. In practice, some data repositories require formal citations provided by the repository when publishing academic papers using the data, while others require acknowledgment of data sources in paper acknowledgments. The former mostly applies to repositories collaborating with journals, where published data are mostly selected, processed resultant datasets, while the latter is more common in data centers under national geoscience research institutions, such as NOAA and USGS, where shared data are mostly large-scale original collection data or fundamental geographic data, with sharing methods including FTP services or REST APIs in addition to data publishing. As shown in Figure 2 [Figure 2: see original paper], only 26% of geoscience repositories registered on the re3data website fully satisfy all five elements.

Figure 2. Statistics of geoscience repositories based on five key elements of data publishing

3.2 Comparison Based on Scientific Data Management Process

Data publishing is closely related to data management in the research process; thus, absolutely independent data publishing or data management platforms do not exist. Among geoscience data repositories on re3data.org, those providing extensive data publishing services mainly fall into three categories: (1) public data publishing platforms; (2) international data sharing organizations or data centers in the geoscience field, such as WDS; and (3) data compilation and sharing platforms built by national-level geoscience research institutions. The authors summarized content related to publishing in the data management process and compared representative data repositories, as shown in Table 3 .

Table 3. Comparison of data publishing and management processes across three repository types

Aspect	Public Data Publishing Platform (Pangaea)	International Data Sharing Organization/Data Center (WDS) (NSIDC, UK National Geoscience Data Centre)	National Geoscience Research Institution Data Compilation
Data Sources	(1) Data submitted by researchers worldwide; (2) Data from geoscience research projects	(1) Data monitored or collected by NASA and NOAA; (2) Data from NSF-funded projects; (3) Other permafrost research data	Various data from UK Natural Environment Research Council-funded projects
Data Submission Process	(1) Submit data via Pangaea ticket system; (2) Author edits data into tables; (3) Data administrator checks parameters and saves to database	(1) Different submission portals based on funder, using web and FTP services; (2) After preliminary description, data editor reviews for publication; (3) If approved, detailed documentation and metadata submitted via FTP	(1) Use unified data collection portal provided by UK National Geoscience Data Centre; (2) Edit metadata according to project specifications and submit datasets; (3) Data administrator reviews and provides DOI registration
PID Registration Standard	DOI	DOI	DOI

Aspect	Public Data Publishing Platform (Pangaea)	International Data Sharing Organization/Data Center (WDS) (NSIDC, UK National Geoscience Data Centre)	National Geoscience Research Institution Data Compilation
Long-term Preservation Format	Tab-delimited text files (ASCII) or Excel preferred; other formats in ZIP	Supports various formats	Professional formats specified: geotechnical/environmental data as *.ags,* *geophysical data as* .las, *.sgy,* .xtf
Data Quality Control	Data parameter checking	Metadata checking; data integrity and consistency; dataset review by project internal experts	Metadata checking
Metadata Standard	Pangaea custom metadata	NASA DIF standard	Custom metadata referencing ISO19115 geographic information metadata standard
Data Citation Standard	DataCite standard	Custom citation standard	Custom citation standard
Open Access Level	Open	Partially open, partially restricted	Tiered access based on project requirements
Data Discovery Tools	PANGAEA Data Warehouse with unified discovery and advanced query/download tools	Unified discovery tool integrated with NASA Earthdata and IceBridge search	Unified discovery tool with SOAP, REST, WMS, WFS, INSPIRE services

Aspect	Public Data Publishing Platform (Pangaea)	International Data Sharing Organization/Data Center (WDS) (NSIDC, UK National Geoscience Data Centre)	National Geoscience Research Institution Data Compilation
Data Interoperability Services		-	Data interoperability services

4. Joint Publication of Data and Papers

Top-tier academic journals such as *Nature* and *Science* have formally proposed requirements for publishing data related to papers and established relevant data deposit and publishing policies. However, actual practices of jointly publishing data and academic papers remain limited. Most involve deep collaboration between publishers and public data repositories. For example, Elsevier' s Science Direct and Scopus have partnered with Pangaea, requiring authors to store data in Pangaea during paper submission and obtain an accessible link before proceeding to academic paper publication, creating a certain degree of interdependence between data and paper publishing. Elsevier' s platform uses Pangaea' s linked data tools to directly access published data. Pangaea' s platform also provides DOIs of Elsevier-published papers that cite the data, forming an interconnection. Joint publishing combines the advantages of the two aforementioned publishing modes, playing an important role in verifying scientific conclusions and establishing complete academic research trajectories. It provides bidirectional transparent access to academic papers and scientific data. Y. Gil et al. discussed future geoscience paper publishing models in an open science environment, proposing that future scientific papers would include multiple forms such as data, software, and reproducible publications, with ideal characteristics for open science and digital scholarship environments: (1) sharing data, software, and other research products in public repositories; (2) using open licenses; (3) providing metadata to describe data, software, and other research products; (4) assigning unique sustainable identifiers to data, software, and other research products; and (5) enabling citation of all mentioned digital resources within articles. Currently, fully integrated publishing of all forms of academic achievements has not been realized, but the publishing and open access of multi-form academic outputs (papers, data, software, other data products) has become a future trend in publishing.

5. Analysis of Key Issues in Data Publishing

5.1 The Role of Three Publishing Modes in Data Sharing

The three publishing modes play different roles in geoscience data sharing. Data journal publishing holds advantages in academic dissemination, with data papers having greater influence than other publishing modes, though data acquisition must be implemented through data repositories that collaborate with or are recommended by journals. Data published through this mode are mostly selected, processed, calculated datasets or data products, and researchers are more concerned with the data itself than data papers. Repository-based publishing can directly integrate into research data management processes, shortening data release cycles and facilitating data acquisition and reuse. The introduction of domestic and international data policies has actively promoted open science and data governance, making research data management necessarily implemented through data repositories. Additionally, large-scale fundamental datasets (such as remote sensing and basic geographic data) can achieve interoperability through repository FTP services or API services. However, most repositories lack academic peer review, with data quality control mostly limited to technical checks and repository certification. Joint publishing combines the advantages of the above two modes, which is crucial for verifying scientific conclusions and establishing complete academic research trajectories, though this requires publishers, libraries, data governance departments, and research institutions to collaboratively establish a knowledge ecosystem.

5.2 Peer Review Issues in Data Publishing

Peer review is critical for data quality control. While not mandatory for data publishing, it represents the gold standard for increasing data credibility. However, a peer review system specifically for scientific data is currently lacking. Most data papers in the publishing field adopt peer review methods similar to academic papers to control data quality. This approach leverages the influence of existing academic journals to drive researchers'recognition of data paper credibility. However, traditional academic papers and data papers focus on different aspects: data publishing emphasizes data reusability, while academic papers emphasize scientific discovery innovativeness. B. Lawrence et al. proposed that scientific data peer review should evaluate three aspects: data quality, metadata quality, and other general factors. ESSD' s two-stage open peer review utilizes an open online period to allow actual data users to judge dataset quality, with user feedback being crucial for verifying consistency between data and paper content and data quality. Additionally, the timing of data review is critical. Qu Baoqiang et al. suggested that with exponential data growth, post-publication peer review may become more common. Post-publication review forms include soliciting opinions on published data, data usage metrics, and data modifications, which may be more scalable for data users. Some repository-based data publishing has also introduced peer review concepts and methods, such as the U.S. National Snow and Ice Data Center, which uses project-internal experts

to review data submitted to the center. Since these experts possess relevant professional knowledge and understand data reuse methods, their review holds authority for data quality control. Data review should differ from traditional academic paper review, focusing not only on academic aspects of data papers but more importantly on data reusability in participating in research and generating reproducible results, metadata quality, and completeness of data usage descriptions. Data peer review experts need to have relevant professional backgrounds and experience using similar data.

5.3 The Importance of Hierarchical Metadata in Geoscience Data Publishing

Metadata primarily describes data to help researchers achieve data reuse. The data quality control investigated in this paper all includes metadata checking. In practical data applications, metadata requires hierarchical description. First, data usage needs can be divided into three levels: data discovery, data citation, and professional data description. The discovery level typically uses Dublin Core metadata standards, the citation level mainly adopts or references DataCite metadata standards, while professional data description is more complex. The most significant characteristic distinguishing geoscience metadata from general metadata is the spatiotemporal nature of the data itself. Commonly used metadata standards in the geoscience field include NASA' s DIF, ISO19115, ISO19139, and FGDC. Most data repositories adopt or reference these metadata standards for professional metadata description, with some even providing multiple standard metadata simultaneously. Second, given the large volume of geoscience data, multi-level metadata should be provided according to dataset granularity. In the investigated data repositories, most specify that individual datasets should not exceed 1-2GB, while original collection data may be much larger. For example, seismic reflection profile data are organized by receivers on the profile, and remote sensing data are gridded by Earth' s latitude and longitude, with entire study areas potentially containing multiple datasets. When Pangaea publishes large-volume reflection seismic data, it decomposes datasets into multiple data sequences, assigning each a unique identifier. This requires providing professional metadata for each data sequence while also providing metadata for the entire study area dataset. In summary, hierarchical metadata description is of great significance for data preservation and reuse in geoscience data publishing practices.

From this research, we can observe that the article volumes of existing geoscience data journals show a yearly increasing trend, indicating that data publishing is gradually being recognized and accepted by researchers, which promotes active data sharing. The introduction of scientific data management policies domestically and internationally has made scientific data preservation and management an essential part of research, and data publishing is closely related to data management. Regardless of the publishing mode, data repositories are indispensable. The authors extracted five essential elements for data publishing, but investiga-

tion reveals that few geoscience repositories registered on re3data fully satisfy these five elements, meaning most repositories have not yet developed comprehensive data publishing capabilities. Joint publishing of academic papers and data is limited by the open science environment and remains rarely practiced. Based on these findings, the authors discuss existing key issues, which provide reference for China' s geoscience field in transitioning from data sharing models to data publishing practices.

Furthermore, geoscience data publishing has its own domain characteristics. The geoscience research process is typically summarized in three stages: data collection through instrumentation, data analysis and processing, and generation of data products and research conclusions through methodological development. Data generated in these three stages include: originally collected data, software for data reading, conversion, and visualization, new data products resulting from novel data processing or computational methods, and integration of multiple data sources and derivatives. Publishing these diverse data products poses a challenge for geoscience data sharing.

References

[1] Chu Y, Zhu Q, Feng Z, et al. Research on the open sharing mechanism of scientific big data and its implications for environmental information sharing[J]. Chinese Journal of Environmental Management, 2015, 7(6):38-45.

[2] Wang J, Sun J. Review, transformation and prospect of World Data Center (WDC)[J]. Advances in Earth Science, 2009, 24(6):612-620.

[3] Chu Y, Sun J, Liao S, et al. Research and practice on earth system science data sharing[J]. Journal of Geo-information Science, 2010, 12(1):1-8.

[4] Wu L, Nan Z, Wang L. Scientific data publishing: A new model to promote data sharing[J]. China Science and Technology Resources Review, 2014(5):72-78.

[5] EAR Division Data Sharing Policy[EB/OL]. [2018-12-18]. https://www.nsf.gov/geo/geo-data-policies/ear/ear-data-policy-apr2018.pdf.

[6] He L, Chang Y. Research progress on scientific data publishing at home and abroad[J]. Library and Information Service, 2014, 58(5):104-110.

[7] Brase J, Farquhar A, Gruttemeier H, et al. Approach for a joint global registration agency for research data[J]. Information Services & Use, 2009, 29(1):13-27.

[8] Kratz J E, Strasser C. Researcher perspectives on publication and peer review of data[J]. PLOS ONE, 2015, 10(2):e0117619.

[9] Li D, Wu Z. Research on research data citation[J]. Library Journal, 2013, 32(5):65-71.

[10] Pronk T E, Wiersma P H, Van Weerden A, et al. A game theoretic analysis of research data sharing[J]. PeerJ, 2015(3):e1242.

[11] WDS[EB/OL]. [2018-12-18]. http://www.icsu-wds.org/organization/intro-to-wds.

[12] Waterman K J, Sierman B. Survey of DSA-certified digital repositories: report on the findings in a survey of all DSA-certified digital repositories on investments in and benefits of acquiring the Data Seal of Approval (DSA)[R]. Hague: Netherlands Coalition for Digital Preservation, 2016:19.

[13] CoreTrustSeal[EB/OL]. [2019-01-12]. https://www.coretrustseal.org/about/.

[14] Huang R, Li N. Research on types of open data license agreements[J]. Library, 2016(8):16-21.

[15] Data Citation Synthesis Group. Joint Declaration of Data Citation Principles[EB/OL]. [2018-02-18]. https://www.force11.org/group/joint-declaration-data-citation-principles-final.

[16] DataCite Metadata Working Group. DataCite Metadata Schema for the Publication and Citation of Research Data[EB/OL]. [2018-02-18]. http://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_{v4}.0.pdf.

[17] Wang D. User testing research on scientific data publishing platforms[J]. Information and Documentation Services, 2017(6):58-63.

[18] National Aeronautics and Space Administration—Global Change Master Directory. Directory Interchange Format (DIF) Writer' s Guide[EB/OL]. [2017-10-15]. http://gcmd.nasa.gov/add/difguide/.

[19] Gil Y, David C H, Demir I, et al. Toward the geoscience paper of the future: best practices for documenting and sharing research from data to software to provenance[J]. Earth and Space Science, 2016, 3(10):388-415.

[20] Lawrence B, Jones C, Matthews B, et al. Citation and peer review of data: moving towards formal data publication[J]. International Journal of Digital Curation, 2011, 6(2):4-37.

[21] Qu B, Wang K. Scientific data peer review from the perspective of data publishing[J]. Library Journal, 2017, 36(10):71-77.

Author Contributions:

Han Lu: Data collection related to the paper, research on key issues in data publishing, and paper writing.

Ding Yi: Research on data repositories and practice of data publishing workflows.

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