Research on Content Standardization of Data Papers (Postprint)

Authors: Huang Guobin, Zheng Xia

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Standardized management, citation, and reuse of scientific data have garnered widespread attention in the academic community. Against this backdrop, data papers and data journals aimed at promoting the rational use of scientific data have emerged in large numbers. However, data papers have not yet developed a unified, standardized format. This study summarizes and distills the content frameworks and core modules of data papers published in currently active data journals in the data publishing community, with the aim of providing references for relevant individuals or institutions in the writing, use, and management of data papers.

[Method/Process] Employing a comprehensive application of content analysis and comparative research methods, and based on the cognitive patterns of scientific data users in searching for, acquiring, and reusing scientific data, this study analyzes the content frameworks and core modules of data papers from six dimensions—topic relevance, data quality, data generation and acquisition methods, application scenarios, usage methods, and supplementary notes—according to the submission guidelines or writing instructions issued by different data journals.

[Results/Conclusion] Currently, no single data journal's submission guidelines or writing instructions for data papers can cover all modules of a data paper; the content composition of data papers is both related to and distinct from that of traditional academic papers; the essential modules of the data paper content framework focus on describing the prominent characteristics of scientific data; a standardized descriptive framework that reflects the characteristics of scientific data has not yet been established.

Full Text

Preamble

Research on Content Standardization of Data Papers

Huang Guobin, Zheng Xia
School of Government, Beijing Normal University, Beijing 100875

Abstract

[**Purpose/Significance**] The standardized management, citation, and reuse of scientific data have attracted widespread attention in academia. In this context, data papers and data journals have emerged in large numbers to promote the rational use of scientific data. However, data papers have not yet developed a unified and standardized format. This study summarizes and refines the content frameworks and core modules of data papers published by currently active data journals, aiming to provide reference for relevant personnel or institutions in the writing, use, and management of data papers. [**Method/Process**] Using comprehensive content analysis and comparative research methods, and based on the thinking patterns of scientific data users when searching for, acquiring, and reusing research data, this study analyzes the content framework and core modules of data papers from six dimensions: subject relevance, data quality, data generation and acquisition methods, application scenarios, usage methods, and supplementary instructions, according to the submission guidelines or writing instructions issued by different data journals. [**Results/Conclusions**] Currently, no single data journal's submission guidelines or writing instructions fully cover all modules of data papers. The content composition of data papers is both related to and distinct from that of traditional academic papers. The essential modules of the data paper content framework focus on describing the prominent characteristics of scientific data. A normalized description framework that reflects the features of scientific data has yet to be established.

Introduction

In the era of big data driven by data, researchers no longer focus solely on research conclusions but also pay attention to the scientific data that serves as the core support for these conclusions [1]. Tracing the generation process of scientific data from a professional perspective and verifying the rationality and rigor of scientific data have become primary considerations for researchers when determining the reliability of a research paper's main conclusions and deciding whether to cite or draw upon that research. At the same time, researchers are no longer limited to using scientific data created by their own teams but have begun to reuse existing scientific data created or provided by others, exploring new approaches and scenarios for utilizing these data from different dimensions [2]. However, for those who reuse scientific data without participating in the data acquisition process, it is often difficult to grasp the core content and quality value of raw scientific data that involves complex factors such as experimental

design, instruments, reagents, research methods, questionnaires, observations, and simulations, and that lacks relevant syntactic, semantic, and pragmatic information support. Therefore, creating scientific data usage documents that introduce and explain the creation, application scenarios, operational essentials, and limitations of scientific data is crucial. Against this background, data papers have gradually emerged and developed.

Currently, there is no consensus in academia on the concept of data papers, but in terms of content and function, data papers typically include a dataset and a descriptive document for that dataset. The descriptive document is a guiding document that introduces and describes how scientific data were obtained, their content composition, application scenarios, and usage methods to facilitate the reading and use of scientific data. To a certain extent, data papers can help researchers, research institutions, and publication platforms more conveniently read and understand the research process, core arguments, and main conclusions when verifying research results, enable integrated research publication platforms to more efficiently describe, reveal, store, organize, retrieve, and disseminate scientific data and other research outputs supported by scientific data, and help data users more accurately and comprehensively determine the quality and value of scientific data they intend to cite or reuse.

To further standardize the format and content of data papers, existing data paper publication platforms have issued submission guidelines, writing instructions, and content templates. Therefore, this study selects representative data journals or platforms and their current normative documents for data papers, empirically analyzes the content composition and core elements of data papers based on the logical thinking of scientific data users when reading and selecting scientific data, summarizes their common characteristics, and aims to better promote the content standardization and efficient use of data papers.

Literature Review

Current research on data papers is relatively limited. Domestic scholars mostly refer to them as "数据论文" (data papers), while foreign scholars use the term "data paper" in relevant studies. Using "数据论文" and "数据期刊" (data journals) as keywords, we searched relevant literature in CNKI and Web of Science. The study found that current research on the content composition of data papers both domestically and internationally focuses on three main aspects:

(1) Brief introductions to the content frameworks of data papers specified by individual data journals. Scholars such as Liu Fenghong et al. [3-4] analyzed the composition of data papers in journals *Data in Brief* and *Ecology*, pointing out that data papers typically include elements such as title, abstract, author, copyright statement, research methods, data files, data records, technical validation, and usage instructions.

(2) Summarizing common content modules of data papers using several data journals as examples without clear sample selection criteria. Tian Ji and

Chen Huixia [5] analyzed content elements of three data journals from perspectives including abstract, background introduction, data methods, dataset, data description, and additional information. Huang Ruhua and Li Nan [6] analyzed the descriptive structure of data papers from the publication policies of seven data publishing journals, concluding that they typically include title, author, abstract, method identifier, storage location, and references. Li Xiu [7] selected seven data journals and analyzed their structural standards from the perspective of quality control in data paper publishing, summarizing six common elements: providing accessible links, providing brief descriptions, author contribution statements, data collection methods and tools, data sharing agreements, and dynamic storage requirements. L. Candela et al. [8] surveyed 15 data journals published by different publishing platforms, indicating that data papers' description of scientific datasets mainly includes data availability, conflicts of interest, coverage, data format, data license, and author contributions. Subsequently, M. Sandra et al. [9] built upon L. Candela's research to analyze the submission guidelines, writing guidance, and recommended templates of these 15 data journals, proposing a general data paper structure containing 11 elements: title, author, author affiliation, abstract, introduction, methods, dataset description, figure/table descriptions, acknowledgments, links, and references.

(3) Research on the components of data journals from perspectives such as the differences between data papers and traditional academic papers, and the mapping between data papers and specialized metadata. Qu Baoqiang and Wang Kai [10] divided the content modules of data papers into two categories based on their differences from traditional academic papers: information related to traditional academic communication (such as title, author, abstract, keywords, etc.) and information related to datasets (basic dataset information and data processing information). V. Chavan et al. [11] studied the mapping relationship between data papers and GBIF (Global Biodiversity Information Facility) metadata from perspectives including title, author, abstract, introduction, coverage, methods, item description, and dataset description.

Analysis of existing domestic and international research on the content composition of data papers reveals that: In terms of research materials, most literature only extracts common elements from data papers published in data journals rather than from submission guidelines and other instructive documents, which may affect the reliability of conclusions since a few data papers may not fully represent a journal's general requirements; From a research perspective, current studies mainly discuss content modules according to their order of appearance in data paper documents rather than analyzing their internal relationships based on the priority of evaluation criteria used by data paper users when selecting scientific data, including subject relevance, quality assessment, application scenarios, and usage methods; In terms of sample selection, although relevant scholars have briefly introduced the core composition of data papers, the sam-

ple selection criteria remain unclear, and the representativeness of data journals and comprehensiveness of disciplinary coverage need to be verified. Based on this, this study selects the top 10 data journals in terms of published data papers as research subjects, analyzes the common characteristics of current mainstream data journals in terms of data paper content frameworks based on the logical process of data users selecting, extracting, reusing, and sharing scientific data, and examines the submission guidelines and writing instructions of these 10 data journals.

Research Methods and Data Sources

This study investigated the Dryad platform [12], which currently hosts the largest number of data journals (134), combined with search results from Web of Science, Elsevier Science Direct, and Wiley Online Library, and extensively surveyed existing data journals. We excluded data journals that did not provide submission guidelines for data papers or whose content descriptions were incomplete. If different data journals provided similar submission guidelines, we only selected the one with the largest number of published data papers. Ultimately, we selected the submission guidelines, writing instructions, and other normative documents from the top 10 data journals that had published at least 30 data papers. These 10 data journals are: *Data in Brief* [13], *Scientific Data* [14], *Ecology* [15], *BMC Research Notes* [16], *Data* [17], *China Scientific Data* [18], *F1000Research* [19], *Geoscience Data Journal* [20], *Journal of Open Psychology Data* [21], and *Ecological Research* [22]. Journal details are shown in Table 1 .

Among these, pure data journals refer to those that exclusively publish data papers, while hybrid data journals publish both data papers and traditional academic papers. Based on this, we employed content analysis and comparative research methods to analyze the content framework and core elements of data papers from six perspectives: subject relevance, data quality, data generation background and acquisition methods, application scenarios, usage methods, and supplementary instructions.

Content Framework and Composition of Data Paper Submission Guidelines

Overall Content Composition

Based on the content characteristics of the submission guidelines and writing instructions from the 10 selected data journals, we designed six primary categories—including subject relevance, data quality, acquisition methods and experimental methods, application scenarios, usage methods, and supplementary instructions—to comprehensively cover all 25 sub-items appearing in these normative documents. Using content analysis, we then statistically analyzed the distribution of each sub-item across the 10 data journals, forming Table 2 "Statistical Analysis of Content Modules in Submission Guidelines of 10 Data Journals." Table 2 shows that these 10 data journals' submission guidelines cover

multiple primary categories, but most have not yet made explicit requirements regarding application scenarios.

Analysis of Six Modules

Data papers are instructive documents that guide the use of scientific data and serve as guides for scientific data users. Following the logical thinking of scientific data users, after directly locating a scientific dataset based on specific paper clues, they typically first consider subject relevance—whether the scientific data is closely related to their research discipline, specialty, direction, field, or topic—followed by data quality, acquisition methods, and experimental methods. For users preparing to reuse scientific data, they will further examine application scenarios, usage methods, and relevant supplementary instructions. Therefore, although different data journals may arrange these modules in different orders and positions due to their editorial styles, comprehensively covering these six basic modules is crucial for data papers to better meet the reading needs of scientific data users.

Subject Relevance Subject relevance reveals the discipline, specialty, direction, field, or topic involved in the scientific data described in a data paper and is the primary consideration for scientific data users. Whether the subject description of scientific data is accurate and sufficient directly affects the orderly organization and efficient retrieval of scientific data by storage institutions and publication platforms, and also directly impacts scientific data users and managers in finding, using, and evaluating scientific data. Among the 10 surveyed data journals, the sub-modules reflecting subject relevance mainly involve the title, abstract, discipline field, dataset, and keywords, as shown in Table 3 .

(1) **Title**. The title specified in data journal submission guidelines refers to a concise statement that summarizes and distills the main content of the data paper, revealing the core information of the data. By quickly browsing the title, researchers can conveniently determine the disciplinary scope and utilization value of the scientific data. Analysis shows that all 10 selected data journals require a title module for data papers. However, Journal 8 does not provide detailed introduction to the title module, while the other nine journals have specific requirements for data paper titles in their submission guidelines.

In terms of content, most data journals require that the title should concisely reflect the data itself, showing the source, temporal, and spatial scope of the data object. In terms of format, titles should minimize or avoid abbreviations or unnecessary punctuation. Beyond these common requirements, different journals have some variations in describing the content, format, and notes for the title module.

(2) **Abstract**. The abstract specified in data journal submission guidelines refers to a brief and accurate short text summarizing the important con-

tent of scientific data, providing further detailed explanation of the title. Overall, the top 10 data journals in terms of published data papers have all provided detailed instructions on abstract content and format. Regarding content, all 10 data journals stipulate that the abstract should only provide brief objective descriptions of data content, data generation background, experimental conditions and methods, and data value, and must not include: abstract derivation or analytical interpretation of scientific data; explanation of new scientific research or discoveries; introduction of research papers associated with the dataset. Regarding format, the submission guidelines of all 10 data journals mention that the abstract should be a standalone paragraph. However, different data journals have varying word count requirements for abstracts (see Table 3). Additionally, unlike other journals, only Journal 10 allows references in the abstract, requiring them to be placed between the abstract and keywords.

(3) **Discipline Scope**. The discipline scope specified in data journal submission guidelines refers to the relatively independent knowledge category to which scientific data belong, used to define and distinguish different research fields involved in the scientific data. Analysis shows that currently only Journal 1 provides standardized definitions of the discipline scope for scientific data described in data papers, requiring information on both primary disciplines (such as physics, chemistry, biology) and secondary disciplines (more detailed disciplinary categories). To provide more efficient retrieval access, Journal 1 also has a "category" system where scientific data providers can select appropriate categories from 27 primary categories and 238 secondary categories when submitting data to facilitate discovery, sharing, and use.

(4) **Dataset Description**. The dataset description specified in data journal submission guidelines refers to detailed descriptions of scientific data content, research background, storage paths, and other supplementary information required from data paper creators during the data description phase. This is one of the differences between data papers and traditional academic papers. Among the top 10 data journals in terms of published data papers, seven have submission guidelines that address dataset description, while Journals 2, 7, and 10 have not made this mandatory (see Table 3). Analysis shows that the seven journals have similar provisions regarding dataset description items, such as requiring dataset title, data type, data content, data generation background, and processing methods. Journals 1 and 3 specifically provide standardized metadata description items, and Journal 6 has added modules such as "Database (Dataset) Basic Information Introduction" to clarify scientific data ownership information.

(5) **Keywords**. Keywords specified in data journal submission guidelines refer to key terms extracted and refined that can cover the full content of the data paper, used for precise and efficient retrieval and location of data papers. Among the 10 selected data journals, except for Journal 2

which does not explicitly require keywords in its submission guidelines, the other nine journals require data papers to provide keywords. Notably, although Journals 1 and 9 require keywords, they do not mandate the number of keywords, while the remaining seven journals explicitly specify the number (see Table 3). Additionally, Journal 6 requires keywords to appropriately reflect temporal and spatial spans.

Data Quality After determining subject relevance, researchers need to evaluate the quality of scientific data to further confirm the reliability of a research paper's experimental design and conclusions, or to decide whether to cite or reuse the scientific data. Analysis shows that most data journals present scientific data quality through five modules: data paper authors, dataset ownership information, dataset completeness, dataset quality control, and funding information, as shown in Table 4 .

(1) **Data Paper Authors**. Data paper authors refer to the main researchers who create data papers. All 10 selected data journals generally require author information including name, email, affiliation, address, and postal code, and designate one author as the corresponding author. Journal 2 has more detailed requirements, allowing up to six authors with equal contributions and six co-supervisors to be listed, with footnotes indicating "These authors contributed equally to this work" or "These authors jointly supervised this work." Journal 5 requires complete author address information following PubMed/MEDLINE standard format (city, postal code, state/province, country, and all email addresses).

(2) **Dataset Ownership Information**. Dataset ownership information specified in data journal submission guidelines refers to ownership information involving all key stages of dataset collection, creation, publication, and utilization. The main modules for describing dataset ownership in these 10 journals include dataset creators, dataset publishers, dataset author institutions, dataset author emails, and dataset collaborators. Basically, all 10 journals require listing dataset creators, collectors, and publishers to varying degrees. Journals 1 and 3 specifically provide standardized metadata description items, and Journal 6 adds modules like "Database (Dataset) Basic Information Introduction" to clarify ownership.

(3) **Dataset Completeness**. Dataset completeness comprehensively describes scientific data from temporal, spatial, type, and format dimensions to reflect the comprehensiveness and continuity of the data. Completeness is a necessary condition for ensuring high-quality data. However, most of the top 10 data journals have not made mandatory requirements for dataset completeness, with only Journal 6 briefly mentioning basic completeness requirements. In its "Database (Dataset) Basic Information Introduction" module, it requires describing dataset composition, such as the number of components and data volume.

(4) **Dataset Quality Control**. Dataset quality control mainly involves usage records, limitations, and quality management of scientific data. The survey reveals: Regarding usage records, only Journal 3 indicates that dataset usage history should be provided, including data request records (recording names of requesters, purposes, and actual usage processes), dataset update records (describing all update operations), and comments and questions from other users (suspicious or anomalous data discovered, limitations encountered, unresolved issues). Regarding dataset limitations, only Journal 4 among the 10 journals requires a description of dataset limitations (within 300 words), including problems during data collection (such as small sample size, outdated data). This journal believes that truthful description of data limitations is crucial for data citation and reuse. Quality control and evaluation. Journals 3, 7, and 9 all require detailed descriptions of data quality control and evaluation, such as identification and handling of outliers, reference standard verification, error data and precision, and missing data points in continuous data.

(5) **Funding Information**. Funding information specified in data journal submission guidelines refers to detailed information about the funding sources for projects that created the datasets, including funding agencies, funding levels, and amounts, which can help data users assess data authority. Generally, research data produced with large-scale investment from higher-level funding agencies is relatively more reliable. Journals 1, 3, 6, and 9 all require data papers to provide funding agency or foundation names, grant numbers, and award numbers.

Data Generation Background, Acquisition Methods, and Experimental Methods Data generation background, acquisition methods, and experimental methods refer to the process records of obtaining and creating scientific data based on specific knowledge backgrounds and research content. All 10 selected data journals address this to varying degrees, as shown in Table 5 .

(1) **Data Generation Background and Purpose**. This module requires summarizing the knowledge background and research objectives of data (dataset) generation. Six data journals require providing background and purpose information, asking for general descriptions of data generation background, data sources, potential use value, and experimental purposes.

(2) **Experimental Design, Data Collection, and Processing Methods**. This module requires describing and introducing various methods and tools for collecting, generating, and processing scientific data, forming the foundation and core of data content description. Given its importance, all 10 data journals mandatorily require descriptive information about data processing, including experimental design, sampling methods, field trials, computer processing (such as data standardization, image feature extraction), chemical reagents and instrument models, and technical validation.

(3) **Data Sample Description**. This module provides general descriptions of sample sources and data structures to highlight the authority and reliability of experimental samples. Four data journals have detailed requirements. For example, Journal 1 requires describing: experimental factors (sample preprocessing); experimental characteristics (brief experimental process description); and data collection location (geographic position and GPS coordinates). Journal 6 requires providing typical dataset samples and describing sample sources and data structures, asking authors to follow principles of representativeness and conciseness when selecting data samples, and similarly requires detailed descriptions of geographic latitude/longitude and regions.

Application Scenarios Data application scenarios refer to descriptions by data paper creators of the main uses, technical validation, and potential value of the scientific data they describe. Content analysis of the submission guidelines and writing instructions from the 10 selected data journals shows that application scenarios are mainly reflected in four modules: data availability, data value, data records, and technical validation. Requirements for these modules are shown in Table 6 .

(1) **Data Availability**. Data availability refers to descriptions of situations where data users can reuse scientific data to initiate or complete research work under specific circumstances, including scientific accessibility, comprehensibility, usage efficiency, and effectiveness. Among the 10 journals, Journals 2, 3, 4, and 7 mandatorily require data availability information. Different journals have different requirements. For example, Journal 2 emphasizes providing source code for programs used in dataset generation and processing—the "code availability" module—explaining how users can access and use this code and what restrictions may apply. It also requires software versions and specific variables or parameters used for generating, testing, or processing the dataset.

(2) **Technical Validation**. Technical validation refers to experiments or analyses provided to support the technical quality of the dataset, which can improve dataset credibility and reliability and help researchers reuse and replicate data. Among the 10 journals, Journals 2, 3, and 7 require technical validation information. Specifically, Journal 2 states that technical validation can be provided in figures or tables to prove data reliability, including experiments validating data collection procedures and statistical analysis of experimental errors. Journals 3 and 7 require all validation information for dataset operations, i.e., how to control data errors or biases and ensure data authenticity through quality validation.

(3) **Data Records**. Data records explain and document all data closely related to a research process, including public repository names where information is stored and overviews of data files and formats. Among the 10 journals, only Journal 2 has detailed content regulations for data records,

requiring each external data record to follow specific citation formats and complete data citations. Data records should be visually presented in tables, indicating research samples, data sources, and experimental operations. If data records contain data from analysis and collection, the output process should also be recorded.

(4) **Data Value**. Data value refers to the utility and value of scientific data's specific attributes and functions in meeting research needs. Five data journals provide different detailed descriptions of the data value module (see Table 5). Common features include highlighting innovation in data sources, processing, and quality control, and explaining dataset value in terms of coverage, processing methods, and potential application value.

Usage Methods The core content of data usage methods is to provide reliable usage instructions for permanent scientific records centered on data, helping data creators submit software packages, code, processing workflows, operational steps, and specific procedures used during data creation to data users, thereby promoting proper application, sharing, or reproduction of scientific data. Analysis of the 10 journals' submission guidelines shows that usage method content mainly appears in four modules: dataset storage location, data structure description, data format/type/timeliness explanation, and data usage instructions, as shown in Table 7 .

(1) **Dataset Storage Location**. This refers to the physical storage location or network storage space of data, such as a specific location, public repository, or third-party repository. To facilitate data dissemination and sharing, data papers should provide hyperlinks to data storage locations associated with the described datasets. Five of the 10 data journals mandatorily require descriptions of dataset storage location (see Table 7). These five journals require providing permanent identifiers (such as DOI or URL) in data papers to promote rapid data acquisition and utilization.

(2) **Data Structure Description**. This reveals details about variables and values in scientific data. Except for Journal 9, all nine selected data journals mention data structure description information in their submission guidelines, explicitly requiring all variables and parameters used to generate, test, or process datasets. Journal 3 has particularly detailed requirements, including: variable characteristics, definitions, units, storage types, lists, ranges, missing values, precision, and format; and missing data, anomalous data, and standard errors.

(3) **Data Format, Type, and Timeliness Explanation**. This module systematically describes key elements to consider when citing or reusing scientific data, including data creation time, first usage time, language, version, data scale, data type, data format, and dataset license. All 10 data journals' submission guidelines explicitly require these elements.

(4) **Data Usage Instructions**. This module requires data paper creators to

introduce scientific data usage methods to better promote citation, analysis, validation, or reuse of scientific data. Five data journals require data usage instructions, with common requirements including data usage methods and recommendations, software or code for data generation, data processing workflows and specific steps, and other supplementary materials. Journal 9 also specifies format requirements, asking for 50-200 words describing data usage and reuse methods.

Supplementary Instructions In addition to the five aforementioned content modules, all 10 surveyed data journals have supplementary instructions in their submission guidelines. Overall, these mainly include four items: supplementary information, author contributions, acknowledgments, and references, as shown in Table 8 .

(1) **Supplementary Information**. This module declares patents, conflicts of interest, notes, abbreviation lists, and related research articles associated with scientific data. Regarding conflict of interest statements, five data journals require public disclosure of interests related to scientific data (see Table 8). Regarding patents, Journal 6 states that if patent achievements were obtained during scientific data creation, they can be noted in this module, though this is not mandatory. Regarding appendices, Journal 5 considers them non-mandatory but useful for supplementing other necessary information crucial for understanding and reproducing scientific experiments. Regarding abbreviations, Journals 1 and 4 require defining all professional vocabulary abbreviations in footnotes on the first page of data papers to ensure consistent use. Regarding footnotes, Journals 1 and 4 require continuous numbering with a separate list at the end, though footnote use should be minimized. Regarding figures, tables, and other elements, this includes descriptions of figures, tables, mathematical formulas, video files, units, and symbols used to summarize data generation and analysis output.

(2) **Author Contributions**. This module details each author's contributions to creating and describing data, providing basis for resolving issues of data attribution, quality, and citation. Six data journals explicitly require listing each author's tasks at the end of data papers. Journal 7 has the most detailed requirements, listing 14 assignable tasks including model construction, data management, formal analysis, investigation, protocol design, and project management.

(3) **Acknowledgments**. This module allows data paper creators to thank individuals who provided support in technical assistance, data compilation, writing guidance, translation, and proofreading. All data journals emphasize providing acknowledgment information, typically requiring concise lists of researchers who participated in data collection or provided other technical assistance.

(4) **References**. Similar to traditional research papers, data papers must also reference existing relevant research using unified formats and numbering. Except for Journals 3 and 10 which do not have mandatory requirements, the other eight journals require listing references in order of appearance in the main text at the end of data papers, with special marking for content containing DOI or URL.

Discussion and Conclusions

The above analysis reveals that the content frameworks of data papers specified by the 10 currently active data journals in the scientific data publishing field have the following characteristics:

(1) **No single data journal's submission guidelines or writing instructions fully cover all modules of data papers**. As shown in Table 1, even *Data in Brief*, which has published over 3,000 data papers, only relatively comprehensively incorporates general modules of data papers but does not explicitly address content describing scientific data completeness, dataset quality assessment, or application scenarios. Moreover, most data journals do not specify disciplinary scope, though in reality, researchers prioritize subject relevance when selecting scientific data or data papers. Due to differences in focus among various data journals, it is indeed difficult to develop a comprehensive module framework.

(2) **The content composition of data papers is both related to and distinct from that of traditional academic papers**. Currently, data paper content can be divided into two parts: one part describes dataset-related items, and the other resembles traditional scientific papers. Dataset-related items are unique to data papers and distinct from traditional academic papers, representing important attributes that enable scientific data to be correctly described, cited, and reused. These items cover data quality assessment, data generation background, acquisition and experimental methods, usage methods, and application scenarios. Items similar to traditional scientific papers mainly include title, author, abstract, keywords, and references. Notably, although most data journal submission guidelines require these traditional elements, their content must typically conform to the characteristics of scientific data and the functional positioning of data papers. For example, some data journals require data paper titles to include terms like "data" or "dataset," while abstracts only need to summarize data sources, experimental methods, and research design without presenting main research findings as required in traditional scientific papers.

(3) **Essential modules of the data paper content framework focus on describing prominent characteristics of scientific data**. Overall, the top 10 data journals in terms of published data papers require comprehensive and detailed descriptions of modules including title, ab-

stract, author, experimental design, acquisition and experimental methods, and usage methods (especially data structure description and data format/type/timeliness explanation). These core modules emphasize describing the characteristics of scientific data. For instance, most data journals require data paper titles to reveal data sources and utilization value; abstracts should focus on summarizing research purposes, data generation background, experimental conditions and methods, and data value, with detailed content placed in the main body—data generation background, acquisition and experimental methods, and usage methods. They also require clear identification of data paper authors (researchers who created the data paper), with individuals who contributed to the data paper but do not meet authorship criteria being acknowledged in the acknowledgments section.

(4) **A normalized description framework reflecting the characteristics of scientific data has yet to be established**. Although current data journals have made beneficial attempts in standardizing data paper content, there is still no unified, standardized description framework that reflects scientific data characteristics. This is mainly reflected in: inconsistent module settings, where different data journals have varying module configurations and lack unified standards; inconsistent content depth, where even for the same module, different journals have different depth requirements; and inconsistent expression formats, where different journals have varying format requirements for the same content. These issues create difficulties for data paper creators and users and hinder the standardized management and efficient use of scientific data.

References

[1] Huang Guobin, Qu Yajie, Wang Shu. Analysis of data management functions of UKDA and ICPSR social science data publishing platforms [J]. Library and Information Service, 2017, 61(21): 40-48.

[2] Huang Guobin, Liu Xinran, Jiang Ying. Analysis of external factors affecting scientific data citation [J]. Digital Library Forum, 2017(6): 2-8.

[3] Liu Fenghong, Zhang Tian. Discussion on emerging academic paper publication types in the context of open science—Research element publishing [J]. Chinese Journal of Scientific and Technical Periodicals, 2014, 25(12): 1451-1456.

[4] Liu Fenghong, Cui Jinzhong, Han Fangqiao, et al. Data paper: An emerging academic paper type in the big data era [J]. Chinese Journal of Scientific and Technical Periodicals, 2017, 28(2): 138-144.

[5] Tian Ji, Chen Huixia. Quantitative analysis and reflection on data journals and data papers [J]. Library Tribune, 2016, 36(3): 42-48.

[6] Huang Ruhua, Li Nan. Research on foreign data journal policies based on the data lifecycle model [J]. Library and Information, 2017(3): 36-42, 108.

[7] Li Xiu. Research on quality control of data journal publishing [J]. Editing Friends, 2017(4): 33-37.

[8] Candela L, Castelli D, Manghi P, et al. Data journals: A survey [J]. Journal of the Association for Information Science and Technology, 2015, 66(9): 1747-1762.

[9] Roa-Martinez SM, Vidotti SAB, Santana RC. Proposed structure of a data paper structure as scientific publication [J/OL]. Revista española de documentación científica, 2017, 40(1): e167. [2019-08-09]. http://dx.doi.org/10.3989/redc.2017.1.1375.

[10] Qu Baoqiang, Wang Kai. The emergence and development of data papers [J]. Library and Information, 2015(5): 1-8.

[11] Chavan V, Penev L. The data paper: A mechanism to incentivize data publishing in biodiversity science [J]. BMC bioinformatics, 2011, 12(6): 2399-2405.

[12] Dryad. Lookup your journal [EB/OL]. [2019-01-07]. https://datadryad.org/pages/journalLookup.

[13] Data in Brief. Guide for authors [EB/OL]. [2019-01-17]. https://www.elsevier.com/journals/data-in-brief/2352-3409/guide-for-authors.

[14] Scientific Data. Submission guidelines [EB/OL]. [2019-01-17]. https://www.nature.com/sdata/publish/sub guidelines.

[15] Ecology. Data papers [EB/OL]. [2019-01-17]. https://esa-journals.onlinelibrary.wiley.com/journal/1939917

[16] BMC Research notes. Submission guidelines [EB/OL]. [2019-01-17]. https://bmcresearchnotes.biomedcentral.com/submission-guidelines/preparing-your-manuscript/data-note.

[17] Data. Instructions for authors [EB/OL]. [2019-01-17]. https://www.mdpi.com/journal/data/instructions.

[18] China Scientific Data. Submission instructions [EB/OL]. [2019-01-17]. http://www.csdata.org/p/static/143/.

[19] F1000research. Article guidelines [EB/OL]. [2019-01-17]. https://f1000research.com/for-authors/article-guidelines/data-notes.

[20] Geoscience data journal. Author guidelines [EB/OL]. [2019-01-17]. https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/about/author-guidelines.

[21] Journal of open psychology data author guidelines [EB/OL]. [2019-01-17]. https://openpsychologydata.metajnl.com/about/submissions/.

[22] Ecological research. Data paper template [EB/OL]. [2019-01-17]. http://www.greynet.org/thegreyjournal/datapapertemplate.html.

Author Contributions: Huang Guobin: Responsible for topic selection, paper writing, revision, and guidance. Zheng Xia: Responsible for material collection and organization, paper writing and submission.

Announcement from Library and Information Service

The First Youth Editorial Board of *Library and Information Service* Officially Established

Library and Information Service is a large academic journal (semimonthly) supervised by the Chinese Academy of Sciences and sponsored by the National Science Library, Chinese Academy of Sciences. With 63 years of development history, the journal is based on the first-level discipline of "Library, Information and Archives Management," integrating theory with practice and academia with application, and has played an indispensable role in disciplinary construction and the development of library, information, and archives undertakings.

To better fulfill its academic and social missions, cultivate more young scholars, and further promote theoretical development and practical innovation in the discipline, this journal plans to establish a Youth Editorial Board.

Responsibilities of Youth Editorial Board Members: (1) Care for, support, and publicize the journal; (2) Write, review, and recommend manuscripts (no fewer than 2 articles per year); (3) Provide suggestions on topics, academic activities, and journal development.

Selection Principles: (1) One member per institution, with appropriate consideration for geography and institutions; (2) Age 原则上 under 40, with appropriate extension for Young Changjiang Scholars; (3) Hold a doctoral degree or associate professor position or above, with outstanding academic achievements.

Based on these principles, the first Youth Editorial Board of *Library and Information Service* is composed as follows:

Editorial Board Director: Lu Wei
Deputy Directors: Cao Gaohui, Zhang Pengyi, Yan Hui
Editorial Board Members: (List of names omitted for brevity)

Library and Information Service Editorial Office
November 2019

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