



Data journals: types of peer review, review criteria, and editorial committee members' positions

Sunkyung Seo, Jihyun Kim

Department of Library and Information Science, Ewha Womans University, Seoul, Korea

Abstract

Purpose: This study analyzed the peer review systems, criteria, and editorial committee structures of data journals, aiming to determine the current state of data peer review and to offer suggestions.

Methods: We analyzed peer review systems and criteria for peer review in nine data journals indexed by Web of Science, as well as the positions of the editorial committee members of the journals. Each data journal's website was initially surveyed, and the editors-in-chief were queried via email about any information not found on the websites. The peer review criteria of the journals were analyzed in terms of data quality, metadata quality, and general quality.

Results: Seven of the nine data journals adopted single-blind and open review peer review methods. The remaining two implemented modified models, such as interactive and community review. In the peer review criteria, there was a shared emphasis on the appropriateness of data production methodology and detailed descriptions. The editorial committees of the journals tended to have subject editors or subject advisory boards, while a few journals included positions with the responsibility of evaluating the technical quality of data.

Conclusion: Creating a community of subject experts and securing various editorial positions for peer review are necessary for data journals to achieve data quality assurance and to promote reuse. New practices will emerge in terms of data peer review models, criteria, and editorial positions, and further research needs to be conducted.

Keywords

Data journals; Data peer review; Editorial positions; Peer review system; Peer review criteria

Received: July 20, 2020
Accepted: July 24, 2020

Correspondence to Jihyun Kim
kim.jh@ewha.ac.kr

ORCID
Sunkyung Seo
<https://orcid.org/0000-0003-0175-2311>
Jihyun Kim
<https://orcid.org/0000-0003-0209-4365>

Introduction

Background/rationale: The importance of research data management and sharing has been emphasized in recent years in a variety of scholarly communities. Data publication has

emerged as part of these discussions, and has drawn considerable attention as a way to provide an incentive for data documentation and sharing [1]. Three primary methods of data publication exist (1) submitting data as supplementary materials to traditional journals, (2) submitting data to a data repository, and (3) publishing the data description as a data paper through data journals [2]. Of these methods, data journals make it possible for researchers generating data sets to publish data papers through peer review, thereby helping authors to be rewarded for their contributions and to earn credit through citations.

The publication process of data journals is similar to that of traditional scholarly journals, and their main process is to distribute data sets via peer review and data repositories [3]. Unlike the peer review of research articles, however, data peer review lacks agreement on consistent criteria or standards, and the understanding and approaches of data peer review vary across disciplines [4-6]. Hence, clear definitions do not exist as to how the processes of traditional peer review can be applied to data, or what should be guaranteed through peer review [1].

Some studies have analyzed peer review processes or criteria in data journals. Lawrence et al. [7] introduced the two-stage peer review procedure adopted by *Earth System Science Data* (ESSD) as a data journal, and proposed a generic data review checklist containing three categories: data quality, metadata quality, and general quality. Hrynaszkiewicz and Shintani [8] explained that the main principles of operating *Scientific Data* (another representative data journal) included credit, reuse, quality, discovery, openness, and service. The journal also specified the criteria of data peer review, including experimental rigor and technical quality, completeness, consistency, and data integrity.

Mayernik et al. [9] examined the data review criteria suggested by traditional scientific journals, data repositories, and data journals. For data journals, they analyzed the review criteria suggested by ESSD, *Geoscience Data Journal*, and *Scientific Data*. All three data journals shared an emphasis on the completeness of the data, detailed descriptions, usefulness, and openness and accessibility.

Objectives: Previous studies have mainly focused on peer review systems and criteria in a small number of data journals. Focusing on nine data journals indexed by Web of Science (WoS), in which the proportion of data papers was over 20%, we investigated the type of peer review, review criteria, and the positions of the editorial committee members of the data journals.

Methods

Ethics statement: This study did not involve human subjects. Neither institutional review board approval nor informed consent was required.

Study design: This was a descriptive study based on journals' policies.

Data sources/measurement: This study focused on data journals indexed in WoS, since these journals tend to be prestigious and to have stable operations. The data journals were assigned "data paper" as a document type, and usually multiple document types can be assigned to one journal. WoS only uses the "article" and "review" document types included in the category of citable items to calculate the impact factor (IF) of each journal. Both "data paper" and "article" are assigned as document types to a single document in the data journals indexed in WoS for calculating data journals' IF. On July 2, 2020, the advanced search function of WoS was used to find the number of data journals to which the document type of "data paper" was assigned. From the 7,362 total results in WoS, 93 data journals were found. Of these 93 data journals, nine data journals (for which the percentage of data papers among all articles was over 20%) were finally selected. When sorting data journals in descending order by the percentage of articles that were data papers, there was a considerable discrepancy between the percentage of the ninth journal (22.24%) and that of the 10th journal (7%). Therefore, a percentage of data papers of 20% was used as a cut-off criterion to select the data journals for this study. To survey the peer review systems, review criteria, and editorial committee structures of the data journals, we analyzed each data journal's data peer review policies. If clarification was needed, journal editors were queried via email, and their responses were incorporated into the analysis. To analyze peer review criteria, the present study utilized the generic data review checklist suggested by Lawrence et al. [7] and the data peer review criteria presented by Carpenter [10]. The review criteria were analyzed in three categories: data quality, metadata quality, and general quality.

Results

Characteristics of the nine target journals

Table 1 presents the journal names, publishers, subjects, IF (2019) values, publishing models, number of data papers, number of articles, and the percentage of articles that were data papers in the nine selected data journals. The percentage of articles that were data papers ranged from approximately 22% to 96%.

Type of peer review

The peer review system types of the nine data journals that actively published data papers was identified, as well as whether the journals provided guidelines. The results are presented in Table 2. *Data in Brief* (Elsevier), *Journal of Open Archaeology Data* (Ubiquity Press), and *Data* (MDPI) adopted a

Table 1. Characteristics of the data journals analyzed

Journal name	Publisher	Subject	Impact factor (2019)	Publishing model	No. of data papers	No. of articles	% (data papers/articles)
<i>Data in Brief</i>	Elsevier	Multidisciplinary sciences	NA	OA	5,044	5,238	96.30
<i>Scientific Data</i>	Springer Nature	Multidisciplinary sciences	5.541	OA	990	1,226	80.75
<i>Human Genome Variation</i>	Springer Nature	Genetics, heredity	NA	OA	78	116	67.24
<i>Earth System Science Data</i>	Copernicus Publications	Geosciences, multidisciplinary	9.197	OA	285	511	55.77
<i>Geoscience Data Journal</i>	Wiley	Geosciences, multidisciplinary	2.714	OA	40	74	54.05
<i>Journal of Open Archaeology Data</i>	Ubiquity Press	Archaeology	NA	OA	11	22	50.00
<i>Data</i>	MDPI	Computer science information systems	NA	OA	134	272	49.26
<i>Gigascience</i>	Oxford University Press	Multidisciplinary sciences	5.993	OA	149	594	25.08
<i>Biodiversity Data Journal</i>	Pensoft	Biodiversity conservation	1.331	OA	125	562	22.24

NA, not available; OA, open access journal.

Table 2. Types of data journal peer review systems

No	Journal name (publisher)	Peer review system form	Peer review guideline
1	<i>Data in Brief</i> (Elsevier)	Single-blind (a minimum of two independent expert reviewers)	Yes
2	<i>Scientific Data</i> (Springer Nature)	Single-blind (an editorial board member chooses one or more referees to evaluate the submission)	Yes
3	<i>Human Genome Variation</i> (Springer Nature)	Single-blind (data report manuscripts may be reviewed by 1 referee only)	-
4	<i>Earth System Science Data</i> (Copernicus Publications)	Interactive two-stage process involving the scientific discussion forum Earth System Science Data Discussions	Yes
5	<i>Geoscience Data Journal</i> (Wiley)	Single-blind	Yes
6	<i>Journal of Open Archaeology Data</i> (Ubiquity Press)	Single-blind	Yes
7	<i>Data</i> (MDPI)	Single-blind Reviewer suggestions (it is possible for authors to suggest three potential reviewers with the appropriate expertise to review the manuscript)	Yes
8	<i>Gigascience</i> (Oxford University Press)	Open review (non-anonymous)	Yes
9	<i>Biodiversity Data Journal</i> (Pensoft)	Community review	Yes

single-blind model for peer review. For *Geoscience Data Journal* (Wiley), no statement was given regarding its peer review system policy. Springer Nature’s *Scientific Data* and *Human Genome Variation* explicitly mentioned “blind review” on their website, but they did not specify whether the process was single- or double-blind. However, upon querying the journal editors, it was discovered that both journals utilized a single-blind model. *Gigascience* (Oxford University Press) adopted an open review model, in which neither authors nor referees remain anonymous.

All peer review models have specific strengths and weaknesses. With the recent emphasis on transparency in the peer

review culture, the data peer review model has evolved into a new modified model from the traditional review approach. ESSD (Copernicus Publications) sought to guarantee the basic scientific and technical quality of the manuscripts that it publishes by carrying out an initial access review by an editor, and then applying an interactive peer review process that supports follow-up interactive discussion and reviews, including public comments from authors and members of the scientific community. *Biodiversity Data Journal* employed a community peer review system that enables experts in various scholarly communities to join peer reviews to distribute the peer review efforts and to enhance transparency and scientific quality.

Table 3. Peer review criteria provided by the data journals

Criteria	<i>Data in Brief</i>	<i>Scientific Data</i>	<i>Earth System Science Data</i>	<i>Geoscience Data Journal</i>	<i>Journal of Open Archaeology Data</i>	<i>Data</i>	<i>Gigascience</i>	<i>Biodiversity Data Journal</i>	Total
Data quality									
Methodological appropriateness	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	7
Acceptable data format	Yes	-	Yes	Yes	Yes	Yes	-	Yes	6
Plausible data values	Yes	-	Yes	Yes	-	-	-	Yes	4
Usefulness of data	Yes	Yes	Yes	Yes	-	-	-	-	4
Identifier of data	-	-	Yes	Yes	-	Yes	Yes	-	4
Sources of errors identified	-	-	Yes	Yes	-	Yes	-	Yes	4
Originality/ novelty of science	-	-	-	Yes	-	Yes	-	Yes	3
Accuracy	-	-	Yes	-	-	Yes	Yes	-	3
Consistency	-	-	Yes	-	-	-	-	Yes	2
Meaningful coverage of data	-	-	-	Yes	-	-	-	Yes	2
Completeness of data	-	-	Yes	-	-	-	-	Yes	2
Validated data	-	-	Yes	-	-	-	-	-	1
Integrity of data	-	Yes	-	-	-	-	-	-	1
Metadata quality									
Sufficient metadata/description of methodologies	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes	7
Accuracy of data description	-	-	-	-	Yes	Yes	Yes	Yes	4
Metadata conforming to standards or template	Yes	-	Yes	-	-	Yes	-	-	3
Completeness of data description	-	Yes	-	Yes	-	-	-	-	2
Metadata about the ownership of data	-	-	-	Yes	-	-	-	-	1
General									
Open license requirement	-	-	-	-	Yes	Yes	Yes	Yes	4
Data availability	-	-	-	Yes	-	Yes	Yes	Yes	4
Complete and appropriate references or acknowledgement	-	-	Yes	Yes	-	-	-	Yes	3
Suitable data repository	-	-	-	-	Yes	-	Yes	Yes	3

When a manuscript is submitted, it is assigned to a subject editor who determines whether the manuscript fits the journal's scope and whether to carry out a peer review. If a peer review is warranted, the subject editor requests two or three "nominated" referees and "panel" referees to conduct a peer review. A nominated referee should complete peer reviews within a given period, whereas a panel referee has no obligation to carry out a peer review. Furthermore, referees can choose to be anonymous or non-anonymous.

Criteria of peer review

All nine data journals, except for *Human Genome Variation*, suggested peer review criteria (Table 3). Among data quality

criteria, a criterion related to methodological appropriateness (e.g., "Are the protocol/references for generating data adequate?" [*Data in Brief*]) was suggested by the largest number of journals. Six journals specified a review criterion related to an acceptable data format (e.g., "The deposited data must include a version that is in an open, non-proprietary format" [*Journal of Open Archaeology Data*]), relating to whether data would be presented in an open format, a common data format, or according to the standards established by scholarly communities. Of the nine data journals, four stated that the data values should be plausible (e.g., "Are the data values physically possible and plausible?" [*Geoscience Data Journal*]), that the data be useful (e.g., "The reuse value of the resulting data-

sets” [*Scientific Data*]), that source of error should be identified (e.g., “Are possible sources of error and noise appropriately described?” [*Data*]), and that the data should have an identifier (e.g., “Is the data set accessible via the given identifier?” [ESSD]) as review criteria. Regarding identifiers, *Data* and *Journal of Open Archaeology Data* presented DOI as an example of a persistent identifier, and *Gigascience* suggested accession number.

As a review criterion related to metadata quality, seven journals suggested that data should have sufficient metadata/methodology descriptions (e.g., “Are methods and materials described in sufficient detail?” [ESSD]). Aside from that, accuracy of data description (e.g., “Do the metadata accurately describe the data?” [*Data*]) was used as a review criterion by four journals. Among the general quality criteria, open license requirements (e.g., “Is the data and software available in the public domain under a Creative Commons license?” [*Gigascience*]) and data availability (e.g., “Does the manuscript properly describe how to access the data?” [*Biodiversity Data Journal*]) were identified as review criteria by four journals.

Editorial committee members’ positions

Editorial committees normally consist of an editor-in-chief and members of the editorial board, who handle peer review and have obligations and rights to reject or accept manuscripts and organize editorial committees. When a data paper is first submitted, the editor-in-chief and editorial committee make a primary judgment regarding the manuscript’s quality, and then the editorial committee contacts referees if desired. Therefore, organizing an editorial committee is as important as conducting peer review. Since data journals handle data papers and data, examining the positions of editorial committee members is essential. The editorial committee members of the nine data journals are shown in Suppl. 1.

Of the nine data journals, seven (excluding *Biodiversity Data Journal* and *Geoscience Data Journal*) have three to 280 advisory/editorial board members. *Scientific Data* has a total of 280 editorial board members assigned to subjects as follows: biological sciences, 154; earth, environment, and ecological sciences, 65; physical sciences, 34; and social sciences, 27. *Biodiversity Data Journal* has 195 subject editors. The advisory/editorial board members are often organized according to subjects and editorial positions, including section/topic/subject editors.

Some defined positions were found with the responsibilities of evaluating data quality and providing information on data curation. For instance, *Gigascience* has one data editor who plays a wide-ranging role in technical quality review of data. The editorial committee of *Gigascience* also incorporates one data scientist, one principal software engineer, and one sys-

tems programmer analyst. *Data* (MDPI) operates a review board (eight members). As examples of relatively new positions relevant for the peer review process, *Journal of Open Archaeology Data* (Ubiquity Press) has one social media editor, and *Human Genome Variation* (Springer Nature) has one variation nomenclature and database editor.

Discussion

The present study found that most of the nine data journals used a single-blind model for data peer review, while *Gigascience* pursued an open review system. To guarantee the transparency and reliability of peer review, ESSD and *Biodiversity Data Journal* adopted modified models, such as interactive review and community review. In interactive review, the members of various scholarly communities can post their opinions, promoting communication between referees and authors. Community review, which follows a traditional peer review method, enables multiple types of referees, such as subject editors and panel reviewers, to join the peer review process and thereby help to distribute peer review efforts. In addition, the open review system allows referees to choose whether to remain anonymous.

Common emphases in the data peer review criteria were the appropriateness of the data production methodology and a detailed description of the methodology. These criteria are particularly important to facilitate research reproducibility and data reuse. The usefulness of the data was considered relevant when evaluating reusability. Likewise, review criteria regarding whether data comply with data standards or formats commonly used in scholarly communities, whether to provide information on open license and data availability, and whether to offer persistent data identifiers emphasized data accessibility for reuse.

Although the composition of the data journals’ editorial committees generally conformed to that of traditional journals’ editorial committees, the editorial committees of data journals tended to include multiple subject editors or some advisory board members with subject-level knowledge. These findings indicate that the editorial committee played a critical role in professionally understanding the data produced in a particular subject field and judging the value and quality of the data. Some, albeit relatively few, data journals incorporated data editors or experts into the editorial committees to evaluate technical data quality and to support appropriate data curation. This relates to the suggestion made by Callaghan et al., who argued in favor of a plan for reducing the data peer review burden that involved letting a data curation expert review the data’s technical quality and a subject expert review the scientific quality through “split[ting] peer review

up into separate phases carried out by different people” [5].

Conclusion: Interactive or community peer review is a new peer review model applied to data journals, which enables members of various scholarly communities to join peer reviews and helps increase peer review transparency and reliability by choosing an open review. Multiple data journals have suggested review criteria, including the appropriateness of the methodology and the need for a detailed description of the methodology. Most journals also specified the need to assess whether the following were provided: acceptable data formats, open licenses, data availability, persistent identifiers, data usefulness, sources of error, and accurate data descriptions. The editorial committees of data journals have subject editors or operate advisory boards including subject matter experts, since the characteristics and value of data are evaluated differently depending on the subject. In addition, some editorial committees include a data editor to evaluate the technical quality of data. Therefore, data journals need to secure subject editors and to establish subject advisory boards. In addition, to enhance the technical quality of data and data curation, assigning relevant positions is increasingly necessary. As data journals continue to develop as new channels of scholarly communication, new practices of peer review will emerge, and further research is necessary on data peer review.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This study was funded by the Korea Institute of Science and Technology Information (KISTI) (contract number: P19032).

Supplementary Material

Supplementary file is available from: <https://doi.org/10.6087/kcse.207>.

Suppl. 1. The composition of editorial committees of the data journals

References

1. Kratz JE, Strasser C. Researcher perspectives on publication and peer review of data. *PLoS One* 2015;10:e0117619. <https://doi.org/10.1371/journal.pone.0117619>
2. Penev L, Mietchen D, Chavan VS, et al. Strategies and guidelines for scholarly publishing of biodiversity data. *Res Ideas Outcomes* 2017;3:e12431. <https://doi.org/10.3897/rio.3.e12431>
3. Austin CC, Bloom T, Dallmeier-Tiessen S, et al. Key components of data publishing: using current best practices to develop a reference model for data publishing. *Int J Digit Libr* 2017;18:77-92. <https://doi.org/10.1007/s00799-016-0178-2>
4. Callaghan S, Murphy F, Tedds J, et al. Processes and procedures for data publication: a case study in the geosciences. *Int J Digit Curation* 2013;8:193-203. <https://doi.org/10.2218/ijdc.v8i1.253>
5. Murphy F. An update on peer review and research data. *Learn Publ* 2016;29:51-3. <https://doi.org/10.1002/leap.1005>
6. Parsons MA, Fox PA. Is data publication the right metaphor? *Data Sci J* 2013;12:WDS32-46. <https://doi.org/10.2481/dsj.WDS-042>
7. Lawrence B, Jones C, Matthews B, Pepler S, Callaghan S. Citation and peer review of data: moving towards formal data publication. *Int J Digit Curation* 2011;6:4-37. <https://doi.org/10.2218/ijdc.v6i2.205>
8. Hrynaszkiewicz I, Shintani Y. Scientific data: an open access and open data publication to facilitate reproducible research. *J Inf Process Manag* 2014;57:629-40. <https://doi.org/10.1241/johokanri.57.629>
9. Mayernik MS, Callaghan S, Leigh R, Tedds J, Worley S. Peer review of datasets: when, why, and how. *Bull Am Meteorol Soc* 2015;96:191-201. <https://doi.org/10.1175/BAMS-D-13-00083.1>
10. Carpenter TA. What constitutes peer review of data: a survey of published peer review guidelines. *arXiv [Preprint]*. 2017 [cited 2020 Jul 20]. Available from: <https://arxiv.org/abs/1704.02236>