An analysis of data paper templates and guidelines: types of contextual information described by data journals

Jihyun Kim

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

Abstract

Purpose: Data papers are a promising genre of scholarly communication, in which research data are described, shared, and published. Rich documentation of data, including adequate contextual information, enhances the potential of data reuse. This study investigated the extent to which the components of data papers specified by journals represented the types of contextual information necessary for data reuse.

Methods: A content analysis of 15 data paper templates/guidelines from 24 data journals indexed by the Web of Science was performed. A coding scheme was developed based on previous studies, consisting of four categories: general data set properties, data production information, repository information, and reuse information.

Results: Only a few types of contextual information were commonly requested by the journals. Except data format information and file names, general data set properties were specified less often than other categories of contextual information. Researchers were frequently asked to provide data production information, such as information on the data collection, data producer, and related project. Repository information focused on data identifiers, while information about repository reputation and curation practices was rarely requested. Reuse information mostly involved advice on the reuse of data and terms of use.

Conclusion: These findings imply that data journals should provide a more standardized set of data paper components to inform reusers of relevant contextual information in a consistent manner. Information about repository reputation and curation could also be provided by data journals to complement the repository information provided by the authors of data papers and to help researchers evaluate the reusability of data.

Keywords

Data papers; Data journals; Data description; Data documentation; Contextual information

Received: January 20, 2020
Accepted: January 27, 2020

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

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

Introduction

Data sharing is an emerging scholarly communication practice that facilitates the progress of science by making data accessible, verifiable, and reproducible [1]. There are several ways of sharing data, including personally exchanging data sets, posting data on researchers' or laboratories' websites, and depositing data sets in repositories. A relatively novel means of releasing data sets is the publication of data papers, which describe how data were collected, processed, and verified, thereby improving data provenance [2]. Data papers are published by data journals, and the publication process is similar to that of conventional journals, in that data papers and data are both peer-reviewed, amended, and publicly accessible under unique identifiers [3]. Since data papers take the form of academic papers and can be cited by primary research articles, credit can be given to data creators [4].

Data papers contain facts about data instead of hypotheses and arguments resulting from an analysis of the data, as commonly presented in traditional research articles [5]. Their primary purpose is thus to explain data sets by providing "information on the what, where, why, how, and who of the data" [6]. The primary advantage of data papers is their rich documentation of data, which is essential for data reuse. A data paper is usually short and consists of an abstract, collection methods, and a description of the relevant data set(s) [7].

However, previous studies have identified a lack of shared templates/guidelines for data papers across data journals. Candela et al. identified 10 classes of data paper components recommended by data journals: availability, competing interests, coverage, format, license, microattribution, project, provenance, quality, and reuse. The authors noted that a unique identifier indicating data availability, such as a DOI or URI, was the only information provided by all the data journals they examined. Less than half of the data journals asked for information on coverage, license, microattribution, project,

and reuse [8]. In addition, Chen explained that data paper templates/guidelines mostly focus on single data sets—that is, on the item level—and only a few provide collection-level descriptions of data, such as multiple data sets or databases. The author suggested that the granularity of research data that a data paper describes should be specified by data journals [9].

The lack of standardization and the problem of granularity in describing data have been discussed in other studies regarding data documentation and metadata [10-12]. Those studies also indicated that documenting an adequate amount of proper contextual information about data would increase the potential for data reuse. The underlying goal of publishing data papers likewise is to enable data reuse, and such papers are expected to address the challenges that elicit "data reuse failure" [13,14]. In this context, the present study examined the types of contextual information that data journals request to be described and determined how common or variable these types of information are across such journals.

This study aimed to identify the components of a data paper as defined by the templates/guidelines of 24 data journals indexed by Web of Science (WoS), with the document type restricted to data papers. The data paper components were mapped onto the types of contextual information suggested by previous studies [15,16]. Therefore, it may be possible to determine the extent to which data papers published in various journals cover the contextual information that researchers need for data reuse and to identify common and unique components across the data journals. The results would help researchers better understand areas for improvement in the guidance provided by data journals for documenting data and in the roles of data journals.

Methods

This study initially identified a broad set of data journals on the basis of 1) two studies [8,9] that conducted a content anal-

Fig. 1. The process of finalizing the selection of the 24 data journals analyzed in this study [8,9].

ysis of data paper templates and/or guidelines and 2) a list of data journals reported by Akers [17]. Candela et al. [8] analyzed 116 data journals from 15 publishers via web-based searches. Chen [9] created an initial list of 93 data journals on the basis of Akers' list and searches on UlrichsWeb and selected 26 data journals from 16 publishers with appropriate consideration of disciplinary domains. Excluding duplicate journals from the two studies resulted in 106 data journals. As previous studies [8,9] suggested, the vast majority of the data journals were mixed journals (i.e., journals publishing any type of paper, including data papers), and pure data journals (i.e., journals publishing only data papers) accounted for only a small proportion.

This study utilized WoS as a tool for selecting the data journals for the analysis. Despite debates over the trustworthiness of journal impact factors generated by WoS, journals indexed by WoS usually maintain a good status because they need to meet the quality criteria set by the database. Of the 106 data journals, 79 were indexed by WoS. The search was restricted to the "data papers" document type in the advanced search function of the database, and 24 data journals were ultimately selected (Fig. 1) [8,9]. Eighteen of those 24 journals overlapped with those examined by either or both of the aforementioned studies (10 by Candela et al. [8], 1 by Chen [9], and 7 by both). The six other journals, namely, *BioInvasions*

Records, *Data*, *Ecological Research*, *Journal of Hymenoptera Research*, *Frontiers in Marine Science*, and *Comparative Cytogenetics*, were also analyzed in this study.

Of the 24 data journals, seven were published by Springer Nature and six by Pensoft (Appendix 1). All the journals published by Pensoft used the same data paper guideline. Among the seven Springer Nature journals, five BioMed Central (BMC) journals shared a single data paper guideline. The remaining data journals provided their own data paper templates and/or guidelines. Therefore, this study collected 15 distinctive data paper templates and/or guidelines for the analysis.

To investigate the contextual information covered by the data papers, I used the types of contextual information suggested by Faniel et al. [15] and Chin and Lansing [16], who elaborated a range of data contexts reflecting the perspectives of data reusers. Chin and Lansing [16] originally proposed various attributes of scientific and social contexts that facilitated data sharing in biological science collaboratories. Four of these contextual attributes are particularly relevant to the scientific context, and therefore were employed for the analysis (Table 1) [8,9,15,16]. I then mapped the types of contextual information onto the data paper components identified by Candela et al. [8] and Chen [9]. This mapping enabled a preliminary assessment of the relationship between data paper components and contextual information and the develop-

Table 1. Mapping between types of contextual information and data paper components

	Types of contextual information		Data paper components	
	Faniel et al. [15]	Chin and Lansing [16]	Candela et al. [8]	Chen [9]
Data production information	Data collection	Experimental properties	Quality, provenance	Collection
	Specimen and artifact	-	Coverage	Coverage
	Data producer	-	Microattribution	Description (file creators), author's contribution
	Data analysis	Analysis and interpretation	-	-
	Missing data	-	-	-
	Research objectives	-	-	-
Repository information	Provenance	Data provenance	Availability	Identifier, relationship
	Repository reputation and history	-	-	-
	Curation and digitization	-	-	-
Data reuse information	Prior reuse	-	Reuse	-
	Advice on reuse	-	Reuse	-
	Terms of use	-	License, competing interests	Copyright, ethical approval, consent to publication, competing interests
-	-	General data set properties	Format	Description (file format, version, creation date)
-	-	-	Project	Funding statement

ment of a coding scheme for the content analysis.

Table 1 showed that not all the types of contextual information suggested by Faniel et al. [15] matched with the data paper components. Specifically, no data paper component identified by the previous studies corresponded to “data analysis,” “missing data,” “research objectives,” “repository reputation and history,” and “curation and digitization”. There was also an inconsistent definition of the term “provenance” between the studies. Candela et al. [8] defined this notion the “methodology leading to the production of the data set”, which is more similar to “data collection” than the definition by Faniel et al. as “sources of the material or traceability”.

“General data set properties,” suggested by Chin and Lansing [16], corresponded to a data paper component relating to the description of data formats, versions, and creation dates. Moreover, “project,” which was mentioned by Candela et al. [8], referring to information about the initiatives within which data are generated, was the only component that did not match any of the types of contextual information. “Funding statement,” which was identified by Chen [9], was also related to the “project” element. Being aware of information about a project and the funding sources that led to data creation would be useful when considering the possibility of data reuse. Thus, this study considered project information to be additional contextual information.

The coding scheme for analyzing the data paper components was largely based on the types of contextual information suggested by Faniel et al. [15]. In addition, the “general data set properties” component, which was noted by Chin and Lansing [16], was added to the coding scheme. The “project” component was included under the “data production in-

formation” category, which was proposed by Faniel et al. [15], since it was a contextual factor relevant to data creation.

Results

The types of contextual information examined in this study were categorized into four groups: general data set properties, data production information, repository information, and data reuse information. Concerning general data set properties, the study explored whether the 15 data paper templates and/or guidelines required the authors to describe the attributes listed in Table 2. Data creation dates, formats, and versions were mentioned by Chen [9], and the remaining properties were identified during the coding process.

Data file name/title was the only property requested by eight of the templates/guidelines (53.3%), and data format description was requested by seven templates (46.7%). The rest of the data set properties were required infrequently by the data journals, and descriptions of data creation dates and languages were rare. Data type was defined differently by the journals; for example, *Journal of Open Archaeology Data* (JOAD) distinguished among primary data, secondary data, processed data, interpretation of data, and final reports, whereas *Data in Brief* (DiB) classified data by type into tables, images, charts, graphs, and figures.

Data production information tended to be requested more frequently by the data journals than general data set properties (Table 3). All the templates and guidelines required information relating to data collection, mainly regarding data collection steps, sampling strategies, and quality control mechanisms. Descriptions of data producers were required by nine

Table 2. General data set properties identified in the data paper templates/guidelines

General data set properties	J _{BMC} ^{a)}	SD	BMC genet	J _{pensoft} ^{b)}	ER	Ecol	GDJ	ESSD	DiB	FiMS	Data	GigaSci	BIR	IJRR	JOAD
Data creation date															√
Data format		√	√		√	√	√		√						√
Data type							√		√			√			√
Data version				√			√								√
File name/title			√	√		√	√			√	√	√			√
File size						√		√				√			
Language				√											√

SD, *Scientific Data*; BMC genet, *BMC Genetics*; ER, *Ecological Research*; Ecol, *Ecology*; GDJ, *Geoscience Data Journal*; ESSD, *Earth System Science Data*; DiB, *Data in Brief*; FiMS, *Frontiers in Marine Science*; GigaSci, *GigaScience*; BIR, *BioInvasion Records*; IJRR, *International Journal of Robotics Research*; JOAD, *Journal of Open Archaeology Data*.

^{a)}Five BMC data journals that used the same guideline were included: *BMC Bioinformatics*, *BMC Genomics*, *BMC Medical Genomics*, *BMC Medical Informatics and Decision Making*, and *BMC Musculoskeletal Disorders*; ^{b)}Six data journals published by Pensoft that used the same guideline were included: *Biodiversity Data Journal*, *Comparative Cytogenetics*, *Journal of Hymenoptera Research*, *Neobiota*, *Phytokeys*, and *Zookeys*.

templates/guidelines, five of which specifically asked for information about the data creators (author list of data sets [GigaScience] or creators [J_{pensoft}, *Geoscience Data Journal*, *Data*, *JOAD*]). The four remaining journals asked for a description of the authors' contributions or information, which possibly corresponded to the role of data creator. The "project" component was mentioned by seven templates/guidelines, and only two of these required overall project descriptions (*Ecology* and J_{pensoft}). The remaining five journals required information about funding.

Specimen and artifact information was requested by five templates/guidelines of biological science, geoscience, and ar-

chaeology journals, and researchers in these disciplines needed such information for data reuse [15]. Temporal, spatial, or taxonomic coverage (J_{pensoft}, *JOAD*), sample availability or location (*Earth System Science Data*, *DiB*), and descriptions of organisms or tissues (*GigaScience*) were identified. Information about data analysis, missing data, and research objectives was also requested by four or five templates/guidelines. Data analysis information involved how data were processed, and missing data information dealt with data anomalies or noise. The research objectives were often expressed as motivations or rationales for collecting the data sets.

In terms of repository information, all but one journal

Table 3. Data production information identified in the data paper templates/guidelines

Data production information	J _{BMC} ^{a)}	SD	BMC genet	J _{pensoft} ^{b)}	ER	Ecol	GDJ	ESSD	DiB	FiMS	Data	GigaSci	BIR	IJRR	JOAD
Data collection	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Specimens and artifacts				√				√	√			√			√
Data producer	√	√	√	√			√	√			√	√			√
Data analysis		√	√			√				√					
Missing data			√		√	√					√			√	
Research objectives		√	√			√	√					√			
Project	√		√	√		√		√				√			√

SD, *Scientific Data*; BMC genet, *BMC Genetics*; ER, *Ecological Research*; Ecol, *Ecology*; GDJ, *Geoscience Data Journal*; ESSD, *Earth System Science Data*; DiB, *Data in Brief*; FiMS, *Frontiers in Marine Science*; GigaSci, *GigaScience*; BIR, *BioInvasion Records*; IJRR, *International Journal of Robotics Research*; JOAD, *Journal of Open Archaeology Data*.

^{a)}Five BMC data journals that used the same guideline were included: *BMC Bioinformatics*, *BMC Genomics*, *BMC Medical Genomics*, *BMC Medical Informatics and Decision Making*, and *BMC Musculoskeletal Disorders*; ^{b)}Six data journals published by Pensoft that used the same guideline were included: *Biodiversity Data Journal*, *Comparative Cytogenetics*, *Journal of Hymenoptera Research*, *Neobiota*, *Phytokeys*, and *Zookeys*.

Table 4. Repository information and data reuse information identified in the data paper templates/guidelines

Types of contextual information	J _{BMC} ^{a)}	SD	BMC genet	J _{pensoft} ^{b)}	ER	Ecol	GDJ	ESSD	DiB	FiMS	Data	GigaSci	BIR	IJRR	JOAD
Repository information															
Provenance		√	√	√	√	√	√	√	√	√	√	√		√	√
Repository reputation and history															
Curation and digitization						√									
Data reuse information															
Prior reuse		√				√	√								
Advice on reuse		√	√				√		√	√	√	√			√
Terms of use		√	√	√	√	√	√	√	√		√	√			√

SD, *Scientific Data*; BMC genet, *BMC Genetics*; ER, *Ecological Research*; Ecol, *Ecology*; GDJ, *Geoscience Data Journal*; ESSD, *Earth System Science Data*; DiB, *Data in Brief*; FiMS, *Frontiers in Marine Science*; GigaSci, *GigaScience*; BIR, *BioInvasion Records*; IJRR, *International Journal of Robotics Research*; JOAD, *Journal of Open Archaeology Data*.

^{a)}Five BMC data journals that used the same guideline were included: *BMC Bioinformatics*, *BMC Genomics*, *BMC Medical Genomics*, *BMC Medical Informatics and Decision Making*, and *BMC Musculoskeletal Disorders*; ^{b)}Six data journals published by Pensoft that used the same guideline were included: *Biodiversity Data Journal*, *Comparative Cytogenetics*, *Journal of Hymenoptera Research*, *Neobiota*, *Phytokeys*, and *Zookeys*.

asked to describe the provenance of the data, indicating the identifier or location of data (Table 4). Data provenance also referred to the relationship of data with other materials, although only one journal (DiB) required a description of any research articles related to the data. None of the data journals asked for descriptions of the repository reputation and history. Regarding curation and digitization, one journal (*Ecology*) required information on archival procedures, including a description of how the data were archived for long-term storage and access.

Data reuse information, mostly concerning advice on reuse and terms of use, was requested (Table 4). In regard to advice on reuse, the authors were required to describe the potential reuse and value of their data for reuse. Data journals required authors to describe the terms of use, primarily relating to competing interests, but several other aspects existed, including ethical approval and consent for participation, consent for publication, license, copyright, and accessibility requirements.

Discussion

The findings revealed common and unique types of contextual information that the data journals requested authors to describe. The most common form of contextual information documented by the journals was data collection methods, followed by data provenance (repository locations and/or data identifiers). More than or almost half of the templates/guidelines identified data file names/titles and data formats as general data set properties, data production information (including data producer and project), and reuse information (including advice on reuse and terms of use). The results are mostly consistent with those of previous studies [8,9]. Yet, Candela et al. [8] mentioned that descriptions of reuse information are often neglected by data journals, but most of the data journals examined in this study addressed the potential reuse of data. In terms of data provenance (indicating the relationship of data to other objects), only one journal (DiB) in this study required information on this relationship, although Chen [9] identified more instances of this information being required.

The types of contextual information that the data journals never or rarely requested included repository information (repository reputation and history, and curation and digitization) and data set properties (data creation dates and languages). In particular, Faniel et al. [15] stated that repository reputation and history are less easily documented since they are more social and relative than other types of context. Two of the data journals examined in this study (*Scientific Data* and *Earth System Science Data*) provided criteria for recommending data repositories, namely, data access conditions and long-term availability [18,19]. The provision of such repository in-

formation by data journals would help reusers understand the trustworthiness of the repositories where certain data are deposited. While data creation dates will help reusers develop sampling frames and identify changes in data creation contexts [20], only one journal asked for this information.

Data production information regarding data analysis, missing data, and research objectives was not specified by the studies of Candela et al. [8] and Chen [9] (Table 1). However, four or five of the templates/guidelines asked for a description of such information. Furthermore, data journals infrequently requested information on data version, file size, and prior reuse, which three of the templates/guidelines mentioned.

Overall, only a small amount of contextual information was commonly requested by the data journals. They tended to focus more on data production information (data collection, data producer, and project) and reuse information (potential reuse and terms of use) than general data set properties or repository information. With the exception of file names and data formats, descriptions of data set properties were generally lacking. Repository information mostly involved unique identifiers of data, but information about repositories' reputation or their curation practices could be provided by data journals to help readers of data papers assess the reusability of data.

In conclusion, the present study examined types of contextual information that data journals asked authors to describe and determined the extent of variation in this information across certain data journals. The primary motivation of publishing data papers is to make data reusable and reproducible, and data papers should provide extensive data documentation that reflects sufficient contextual information. This study suggests that data journals should provide a more standardized set of data paper components to inform reusers of the various types of context in a consistent manner. Furthermore, data journals should not only require data availability information, but also provide details about the quality of data repositories that would complement the repository information described by data paper authors.

Conflict of Interest

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

References

1. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. *Res Integr Peer Rev* 2017;2:4. <https://doi.org/10.1186/s41073-017-0028-9>
2. Pasquetto IV, Randles BM, Borgman CL. On the reuse of

- scientific data. *Data Sci J* 2017;16:8. <http://doi.org/10.5334/dsj-2017-008>
3. Gray S. Case study: publishing a data paper [Internet]. [place unknown]: DOCZZ; 2015 [cited 2020 Jan 15]. Available from: <http://doczz.net/doc/6729472/case-study%E2%80%94publishing-a-data-paper---research-data-service>
4. Jefferies N, Murphy F, Ranganathan A, Murray H. Data-2paper: giving researchers credit for their data. *Publications* 2019;7:36. <https://doi.org/10.3390/publications7020036>
5. Chavan V, Penev L. The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 2011;12 Suppl 15:S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>
6. Callaghan S, Donegan S, Pepler S, et al. Making data a first class scientific output: data citation and publication by NERC's environmental data centres. *Int J Digit Curation* 2012;7:107-13. <http://www.ijdc.net/article/view/208>
7. Kratz J, Strasser C. Data publication consensus and controversies. *F1000Research* 2014;3:94. <https://doi.org/10.12688/f1000research.3979.3>
8. Candela L, Castelli D, Manghi P, Tani A. Data journals: a survey. *J Assoc Inf Sci Technol* 2015;66:1747-62. <https://doi.org/10.1002/asi.23358>
9. Chen YN. An analysis of characteristics and structures embedded in data papers: a preliminary study. *Libellarium* 2017 Mar [Epub]. <http://dx.doi.org/10.15291/libellarium.v9i2.266>
10. Atici L, Kansa SW, Lev-Tov J, Kansa EC. Other people's data: a demonstration of the imperative of publishing primary data. *J Archaeol Method Theory* 2013;20:663-81. <https://doi.org/10.1007/s10816-012-9132-9>
11. Friedhoff S, Meier zu Verl C, Pietsch C, et al. Social research data: documentation, management, and technical implementation within the SFB 882 [Internet]. Bielefeld: University of Bielefeld; 2013 [cited 2020 Jan 15]. Available from: https://pub.uni-bielefeld.de/download/2560035/2560036/SFB_882_WP_0016_Friedhoff_Meier-zu-Verl_Pietsch_Meyer_Vompras_Liebig.pdf
12. Kim J, Yakel E, Faniel IM. Exposing standardization and consistency issues in repository metadata requirements for data deposition. *Coll Res Libr* 2019;80:6. <https://doi.org/10.5860/crl.80.6.843>
13. Rees J. Recommendations for independent scholarly publication of data sets. San Francisco, CA: Creative Commons; 2010.
14. Costello MJ, Michener WK, Gahegan M, Zhang ZQ, Bourne PE. Biodiversity data should be published, cited, and peer reviewed. *Trends Ecol Evol* 2013;28:454-61. <https://doi.org/10.1016/j.tree.2013.05.002>
15. Faniel IM, Frank RD, Yakel E. Context from the data reuser's point of view. *J Doc* 2019;75:6:1274-97. <https://doi.org/10.1108/JD-08-2018-0133>
16. Chin G, Lansing CS. Capturing and supporting contexts for scientific data sharing via the biological sciences laboratory. Paper presented at: 2004 ACM Conference on Computer Supported Cooperative Work; 2004 Nov. <https://doi.org/10.1145/1031607.1031677>
17. Akers K. A growing list of data journals [Internet]. [place unknown]: Data@MLibrary; 2014 [cited 2020 Jan 20]. Available from: <https://mlibrarydata.wordpress.com/2014/05/09/data-journals/>
18. Earth System Science Data. Repository criteria [Internet]. Göttingen: Copernicus Publications [cited 2020 Jan 20]. Available from: https://www.earth-system-science-data.net/for_authors/repository_criteria.html
19. Scientific Data. Suggesting additional repositories [Internet]. [place unknown]: Nature Research [cited 2020 Jan 20]. Available from: <https://www.nature.com/sdata/policies/data-policies#repo-suggest>
20. Zimmerman A. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *Int J Digit Lib* 2007;7:5-16. <https://doi.org/10.1007/s00799-007-0015-8>

Appendix 1. Twenty-four data journals selected for the analysis

Publisher	Data journal (full name)	Subject area	Publishing model	Pure vs. mixed
Springer Nature	<i>BMC Bioinformatics</i> ^{a)}	Biochemical research methods	OA	Mixed
	<i>BMC Genetics</i>	Genetics, heredity	OA	Mixed
	<i>BMC Genomics</i> ^{a)}	Biotechnology, applied microbiology	OA	Mixed
	<i>BMC Medical Genomics</i> ^{a)}	Genetics, heredity	OA	Mixed
	<i>BMC Medical Informatics and Decision Making</i> ^{a)}	Medical informatics	OA	Mixed
	<i>BMC Musculoskeletal Disorders</i> ^{a)}	Orthopedics	OA	Mixed
	<i>Scientific Data</i>	Multidisciplinary sciences	OA	Pure
Pensoft ^{b)}	<i>Biodiversity Data Journal</i>	Biodiversity conservation	OA	Mixed
	<i>Comparative Cytogenetics</i>	Genetics, heredity	OA	Mixed
	<i>Journal of Hymenoptera Research</i>	Entomology	OA	Mixed
	<i>Neobiota</i>	Biodiversity conservation	OA	Mixed
	<i>Phytokeys</i>	Plant sciences	OA	Mixed
	<i>Zookeys</i>	Zoology	OA	Mixed
Wiley	<i>Ecological Research</i>	Ecology	Hybrid	Mixed
	<i>Ecology</i>	Ecology	Hybrid	Mixed
	<i>Geoscience Data Journal</i>	Geosciences, multidisciplinary	OA	Pure
Copernicus Publications	<i>Earth System Science Data</i>	Geosciences, multidisciplinary	OA	Pure
Elsevier	<i>Data in Brief</i>	Multidisciplinary sciences	OA	Pure
Frontiers Media S.A.	<i>Frontiers in Marine Science</i>	Environmental sciences	OA	Mixed
MDPI	<i>Data</i>	Computer science information systems	OA	Mixed
Oxford University Press	<i>GigaScience</i>	Multidisciplinary sciences	OA	Pure
REABIC Journals	<i>BioInvasions Records</i>	Biodiversity conservation	OA	Mixed
SAGE Publications	<i>International Journal of Robotics Research</i>	Robotics	Subscribe	Mixed
Ubiquity Press	<i>Journal of Open Archaeology Data</i>	Archaeology	OA	Pure

OA, open access.

^{a)}These journals share a single data paper guideline; ^{b)}The journals published by Pensoft use the same data paper guideline.