Korean researchers' motivations for publishing in data journals and the usefulness of their data: a qualitative study

Jungyeoun Lee¹, Jihyun Kim²

¹Ewha Institute of Social Sciences, ²Department of Library and Information Science, Ewha Womans University, Seoul, Korea

Abstract

Purpose: This study investigated the usefulness and limitations of data journals by analyzing motivations for submission, review and publication processes according to researchers with experience publishing in data journals.

Methods: Among 79 data journals indexed in Web of Science, we selected four data journals where data papers accounted for more than 20% of the publication volume and whose corresponding authors belonged to South Korean research institutes. A qualitative analysis was conducted of the subjective experiences of seven corresponding authors who agreed to participate in interviews. To analyze interview transcriptions, clusters were created by restructuring the theme nodes using Nvivo 12.

Results: The most important element of data journals to researchers was their usefulness for obtaining credit for research performance. Since the data in repositories linked to data papers are screened using journals' review processes, the validity, accuracy, reusability, and reliability of data are ensured. In addition, data journals provide a basis for data sharing using repositories and data-centered follow-up research using citations and offer detailed descriptions of data.

Conclusion: Data journals play a leading role in data-centered research. Data papers are recognized as research achievements through citations in the same way as research papers published in conventional journals, but there was also a perception that it is difficult to attain a similar level of academic recognition with data papers as with research papers. However, researchers highly valued the usefulness of data journals, and data journals should thus be developed into new academic communication channels that enhance data sharing and reuse.

Keywords

Information dissemination; Data journal; Data paper; Motivation; Publishing

Received: July 17, 2021

Accepted: July 25, 2021

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

ORCID

Jungyeoun Lee
<https://orcid.org/0000-0003-2543-959X>
Jihyun Kim
<https://orcid.org/0000-0003-0209-4365>

Introduction

Background/rationale: With the development of infrastructure capable of processing large-capacity data, the integration and analysis of data from different disciplines have brought about remarkable scientific advances. The relaxation of restrictions on proprietary scientific data has led to the identification of connections between previously hidden scientific patterns through the revitalization of new data-driven approaches and advanced collaboration [1]. Initially, re-

searchers expressed concern about the lack of incentives for data sharing and did not actively participate in it; however, their interest has increased as research funding agencies such as the US National Science Foundation (NSF) publicly announced policies related to data management and sharing public research results [2].

Until now, academic journals that share processes and analysis results related to research topics have been at the center of academic communication in the field of science and technology. However, with the recent emphasis on the importance of data sharing and reuse, data journals have emerged as a new channel for this purpose. Data journals publish data papers that describe facts about data, such as data collection methods and data features, and the described data are disclosed and maintained in data repositories [3]. In data journals, data and data papers are shared in a citable format through a peer-reviewed quality assurance process so that they can be recognized as research achievements [4,5]. In this respect, data journals have emerged as a new medium for sharing and managing data. Data journals must consider variety of factors, such as the context of research data collection, the description of data collection, and the establishment of infrastructure for the organization, verification, preservation, and reuse of data. In addition, standardization of technology related to data sharing should be a prerequisite. Data journals share an emphasis on the appropriateness of data production methodology and detailed descriptions during the peer review process [6]. Data journals ask authors to provide information on aspects of data production, such as the data collection, data producers and related projects, and data identifiers [7].

The publication of data journals and related research initiatives are actively underway, primarily by publishers and academic societies [4]. As the open science movement has emerged in the research environment, research data have received more attention. For scientific integration and reproducibility, research data have begun to be shared more frequently [8]. This phenomenon has also increased the value of data journals. However, very few studies have investigated the perceptions or experiences related to data journals from researchers' perspectives [9].

Objectives: The purpose of this study was to elucidate the usefulness and limitations of data journals. Qualitative exploratory research was conducted on motivations for submission, review and publication processes, data sharing, obstacles, and differences from existing academic journals according to researchers with experience publishing in data journals.

Methods

Ethics statement: The interview data collected in the study

were recorded with consent in compliance with research ethics concerning personal information protection. The collected data were used for research purposes only, and voice recordings were converted into transcripts and used as basic data for this study.

Study design: This qualitative study was conducted to examine researchers' motivations for, and experiences with, submitting to data journals. The study was described according to the SRQR (Standards for Reporting Qualitative Research) guideline [10].

Researcher characteristics and reflexivity: The researchers are experts in library and information science with more than 15 years of research experience.

Context: Interviews in the form of questions and answers were conducted based on a semi-structured questionnaire (Appendix 1), and the data were analyzed using semantic unit coding and clustering.

Sampling strategy: In order to evaluate the representativeness of researchers who submitted data papers, the 79 data journals indexed in Web of Science were screened to find potential research subjects. Among them, we selected four journals in which data papers accounted for 20% or more of their publication volume (*Data in Brief* 94.5%, *Scientific Data* 77.9%, *Data* 44.3%, and *GigaScience Data* 22.17%) with corresponding authors affiliated with South Korean research institutes.

Data collection methods: Emails were sent to a total of 98 corresponding authors from July 24 to October 15, 2019, and a total of seven research subjects were selected for interviews after three rounds of correspondence. The interview questionnaire consisted of five items (Appendix 1) related to their motivations for publishing a data paper, the necessity of data papers, obstacles related to data paper publication, data sharing, and the possibility of founding a data journal in Korea.

Data collection instrument and technologies: Face-to-face and telephone interviews were conducted, with each interview lasting for an average of approximately 58 minutes (Table 1).

Units of the study: As presented in Table 1, the seven subjects included four university professors and three researchers from governmental research institutes. From the interviews, it was found that they all held PhD degrees and had served as reviewers for international journals. All the participants except one were male, and most interviewees had conducted research in fields related to biological sciences or medicine, such as immunology, medical engineering, bioinformatics, or biochemistry.

Data processing: The contents of the interviews were converted into a transcript, and responses were categorized by theme using Nvivo 12, shown in Fig. 1.

Data analysis: Content analysis was performed by creating group clustering while coding for restructuring relevant

Table 1. Background of interviewees, interview modality, and interview duration

ID	Affiliation type	Position	Gender	Research disciplines	Interview modality	Interview duration
P1	University	Associate professor	Man	Immunology of infection	Telephone	38 min
P2	University	Assistant professor	Man	Medical engineering and technology	In-person	1 hr 17 min
P3	Governmental research institute	Researcher	Man	Biomedical informatics	In-person	1 hr 17 min
P4	University	Assistant professor	Man	Electrical and electronic engineering	In-person	54 min
P5	University	Full professor	Man	Biological engineering	In-person	56 min
P6	Governmental research institute	Senior researcher	Man	Biology	In-person	1 hr 2 min
P7	Governmental research institute	Director	Woman	Biochemistry	Telephone	41 min

<div> <div>Quick Access</div> <div> Files Memos Nodes </div> </div> <div> <div>Data</div> <div> Files File Classifications Externals </div> </div> <div> <div>Codes</div> <div> Nodes </div> </div>	<div> <div>데이터의 활용적 가치</div> <div>Search Project</div> </div> <table> <thead> <tr> <th>Name</th><th>Files</th><th>Referen</th></tr> </thead> <tbody> <tr> <td>데이터 활용</td><td>5</td><td>7</td></tr> <tr> <td>데이터의 독창성</td><td>5</td><td>16</td></tr> <tr> <td>데이터의 신뢰성</td><td>5</td><td>13</td></tr> <tr> <td>데이터의 신속성</td><td>1</td><td>2</td></tr> <tr> <td>데이터의 재현성</td><td>3</td><td>5</td></tr> <tr> <td>우선순위</td><td>1</td><td>1</td></tr> <tr> <td>최신 연구동향 파악 가능_데이터저널</td><td>1</td><td>2</td></tr> </tbody> </table> <div>Drag selection here to code to a new node</div>	Name	Files	Referen	데이터 활용	5	7	데이터의 독창성	5	16	데이터의 신뢰성	5	13	데이터의 신속성	1	2	데이터의 재현성	3	5	우선순위	1	1	최신 연구동향 파악 가능_데이터저널	1	2	<div> <div>데이터의</div> <div>데이터의</div> </div> <div> <div>Reference 1 - 0.72% Coverage</div> <p>일반 프로세스를 그대로 따라가고요. 가장 큰 거는 어떻게 보면 이제 그 일반 저널들은 연구 가치적인 측면에서의 저평가 흥미가 오리지널리티 뭔가 새로운 게 있어야 되잖아요. 그런데 이 데이터 쪽은 기존의 오픈 되어 있는 데이터 대비 새로운 게 있어야 된다는 게 있죠. 또 하나의 오리지널리티가 있어야 좀 더 아무래도 리뷰어들이 봤을 때도 이미 이런 데이터 있는데 굳이 이렇게 얘기를 할 수 있으니까요.</p> <div>Reference 2 - 0.46% Coverage</div> <p>그 똑같은 데이터를 다시 여기 오픈하는 거는 이런 쪽에서도 아마 좀 꺼려하는 거 같더라고요. 그니까 데이터가 모았으면 그 데이터를 완전 오리지널리티로 가지고 어디에도 publication 안 된 원천 데이터를 딱 오픈해줘야 그쪽에서 좀 잘 받아주는 거 같고요.</p> <div>Reference 1 - 1.30% Coverage</div> <p>물결이 있느냐 없느냐를 저희가 타겟으로 하는 건 관상용 풍양귀비라고 하고 한국 국명으로서는 개양귀비라는 표현인데 개양귀비가 과연 실제 마약을 양귀비하고 앞에 댔던 양귀비하고의 성분적인 차이가 얼마나 있고 만약에 아편 양귀비는 국내에 대부분 할 수 없는 제한적인데 여기에도 유용한 물질들이 있으면 의학 소재로 활용 가능한 물질이 있으면 여기에서 한 번 더 시도를 해보자 왜냐하면 국내 한 8군데 정도에 양귀비 축제가 열리고 있고 그런 것들이 순수하게 관상용으로 보고 팔라는 형태로 가거든요. 근데 예전에 아실지 모르겠지만 예전에 뭐 할머니 또래들은 분들이 할머니 분들은 그 배가 아프면 약으로 쓰고 뭐 그런 그렇게 썼거든요.</p> </div>
Name	Files	Referen																								
데이터 활용	5	7																								
데이터의 독창성	5	16																								
데이터의 신뢰성	5	13																								
데이터의 신속성	1	2																								
데이터의 재현성	3	5																								
우선순위	1	1																								
최신 연구동향 파악 가능_데이터저널	1	2																								

Fig. 1. Data coding and clustering process using Nvivo 12.

theme nodes. In order to evaluate the reliability and validity of the study, cross-analysis between researchers was performed. Based on the results of coding performed by two coders for 15% of the total interview data, intercoder reliability was measured using Cohen's kappa and was found to be 0.718, which is within the range of substantial reliability [11]. **Techniques to enhance trustworthiness:** No further process was implemented.

Results

Synthesis

The value of data

Researchers who contribute to data journals support the economic, practical, and educational values of data. The tremendous budgets dedicated to research result in large swaths of data such as original data sets and image data. The high degree of economic investment allocated for the discovery of rare resources as well as the production, collection, and analy-

sis of original data should be shared and used as data sets, and further scientific progress should be made through interdisciplinary research. Research data can be reused through accurate interpretations and analyses. To this end, in addition to an explanation of the data, data quality and data standardization must be considered, and data verification must be performed. Since data journals verify the reliability, validity, accuracy, and reproducibility of data through the review process, the reusability of data sets is increased. As such, data journals curate verified data, produce detailed descriptions of the data collection process and experimental methods in data papers, and, at the same time, provide data free of charge so that other researchers can use them universally. As such, data papers that contain detailed descriptions of research data and reproducible experimental methods have intellectual value since they can set precedents with regard to protocols related to experimental data for use by subsequent researchers (Fig. 1).

The value of data journals as research achievements

Disclosure of original and valuable research data to the academic community became possible because data produced by researchers using data journals are considered academic research achievements for the purposes of annual performance evaluations. Unconditional disclosure of data in the highly competitive science and technology field can be undesirable since researchers lose their monopoly power over data and research results. Important and original data derived from the research process are typically intended to be disclosed after the research achievements are recognized.

I plan to conduct follow-up research with this data, but if I disclose the data, others can proceed with research. Then the research will no longer be as valuable for me. (P2)

Data journals are recognized as academic achievements in the same way as conventional academic journals. The submission and review processes of data papers are not different from those of other types of academic papers, and special emphasis should be given to the handling and processing of the results of data collection.

The reasons why the researchers submitted to data journals were as follows: first, they submitted to conventional academic journals but received a recommendation from reviewers to submit their manuscripts as data papers; and second, they submitted data papers after learning that the editorial committee would publish a special edition or a data note section.

There were a few parts that were slightly unorganized, so after being told to submit a data note rather than an article, I organized them and submitted it all at once. A reviewer recommended it. (P6)

Thus, the biggest motivation for researchers to submit to data journals was to receive recognition for their research achievements. In particular, researchers in South Korea are pressured to publish their research results in a short period of time in journals with high impact factors due to annual appraisals conducted by universities and research institutes. Data journals increasingly present an alternative for receiving academic credit since they allow researchers to present review results and draw research conclusions relatively quickly. However, researchers considered it somewhat disappointing that more significant research achievements could not be obtained using this method, since it left less time for analysis and discussion.

It could have been published in a high-ranking journal if the data had been analyzed well, but I had to produce achievements quickly, which was one of the reasons I chose the data journal. (P6)

The researchers were also concerned that, even if the impact factor of a data journal was high, the paper could be disparaged as presenting results of no research value and viewed as “just a collection of data.” Although data papers are valu-

able in that they promote data sharing and utilization, it was also pointed out that it is not likely for them to be recognized as an achievement that replaces traditional research papers published in existing academic journals. In particular, for researchers with master’s or doctoral degrees applying for research positions, it was considered desirable to prove their research ability with research papers published in academic journals and present data papers as supplementary research achievements.

That’s what I tell my students. This is a data paper, not a research paper, so it may be considered valuable later, but it can be a bit of an issue if you put this as a representative performance, for example, and do something with it later just because the impact factor of a data journal is high. (P2)

The quality and specialization of data journals are gradually being improved, and data journals that initially received a wide variety of data are gradually beginning to favor more meaningful data. In addition, in the case of fields with rapid technological development, data are gathered specifically for data papers and researchers tend to be aware of the latest research trends related to data papers.

In the case of GigaScience, various data were received, but these days they do not accept any data that are not meaningful. (P6)

Repository for data sharing and preservation

When submitting to data journals, authors submit data to data repositories at the same time as they submit data papers. Repositories aim to standardize data, build architecture and infrastructure for the data, and share data [2]. Each publisher has different policies for repositories; publishers’ internal repositories or general-purpose or specific subject area repositories are recommended. Since data in repositories becomes openly available at the time of publication, repositories are essential components of infrastructure for data sharing and preservation related to data journals.

Traditional academic journals also recommend sharing research analysis data in data repositories during the review process, but since the data are not reviewed before being added to the repository, they are often added in an unorganized state. However, since data journals focus on the value of the originality of data, the data collection process, the analysis method, and the usability of data, there are clear guidelines for data standardization, data quality, and data sharing, and data sharing and management in the context of repositories are undertaken in accordance with standards and procedures for their preservation and utilization.

Recently, overseas research funders such as the National Institute of Health and NSF have established a policy that general academic papers with state-funded research results must be deposited in a public access-compliant repository design-

nated by the National Institute of Health or NSF; and the data must be disclosed in a repository [12]. In addition, as researchers have begun to acknowledge the qualitative limitations of research that is limited to individual research fields, voluntary data sharing for facilitating interdisciplinary research has become more common. A high quantity of research data is available in general repositories such as Github, the National Center for Biotechnology Information (NCBI), and institutions' websites. However, it is difficult for researchers to use these data since they are often in the form of large sets of raw data with no detailed descriptions. Thus, researchers often encounter errors in the process of downloading and analyzing these data due to the absence of data standardization and reproducibility verification.

Data papers have data that went through basic analyses. But NCBI often receives just raw data. So, many general users feel at a loss with the data they receive from NCBI. There are even cases where general users cannot perform analyses when it is necessary to do so by themselves. (P6)

As such, there is a difference between the reliability and accuracy of data in repositories linked to data journals and data in general repositories. Therefore, descriptions of the characteristics and collection process of data in data papers increase their reusability, reliability, and utility. Data in data journals have high validity and reliability, allowing for easy utilization. In particular, since it is difficult to obtain high-quality data from experimental studies on humans or animals due to the influence of environmental and technological factors and variables, descriptions of the data collection process are important, and for special data on animals and plants collected directly in remote areas such as Antarctica, the data resource itself is very valuable.

Traditional journals focus on the value of research, whereas data journals require something new compared to existing open data. Reviewers look for original data that have not been published anywhere else. (P2)

The difficulty of routine management and preservation of research data produced in a laboratory setting also leads researchers to consider submitting data papers. This is due to the assurance that the data will be permanently preserved and available in a repository at the same time as the paper is submitted to a data journal.

Although students try to organize their data well, it is not easy to keep archiving data consistently. You have to keep the URL, but after 2 to 3 years, it is difficult. We found that the location of the data kept changing later on. It occurred to me that one of the easiest ways to maintain and share data was to submit a paper to a data journal. (P4)

For operating a repository, data structures must be managed according to the academic field, and metadata, useful

and usable data, and data stability must be maintained. For data sharing and utilization, the content and quality of data should be routinely managed to prevent repositories from simply becoming data containers.

Dissemination of research achievements

You can also get credit since data papers are considered academic achievements. In practice, we believe that citations are more valuable than the credit itself in the long run. (P4)

The value of published academic achievements is confirmed by citations in other studies. The same is true for data journals; academic achievement is confirmed when a data paper is cited and the data are used in other studies.

In fact, after publishing data papers, the researchers received many data-related inquiries via email, with communications ongoing. This phenomenon can also lead to joint research in the future. Ultimately, it was found that data journals act as a channel for academic communication between researchers, going beyond the preservation of data alone to facilitate collaboration with other fields.

The data from the Genome Project are highly versatile. Analyzing the genome of a new organism, such as the human genome, does not end with my research in my laboratory. Various groups working on it use the data to experiment according to their interests. I provide my research as a sort of reference for that kind of thing. (P5)

Discussion

Key results: The results of analyzing researchers' opinions on their motivations for submission to data journals and the usefulness of data are presented in Fig. 2. The most important value of data journals to researchers was to obtain credit for their research achievements using data. In particular, the main motivation for Korea-based researchers to submit to data journals was found to be to quickly publish data papers in data journals with high impact factors to be recognized as research achievements since researchers in Korea receive yearly performance evaluations based on journals' impact factors. However, since data papers describe data without an in-depth discussion, the fact that it is difficult to obtain the same degree of academic recognition with data papers as it is with research papers published in traditional journals was recognized as a limitation.

Interpretation: Unlike traditional journal research papers that strictly distinguish between data and discussion/analysis, data papers obscure such distinctions. Thus, there is controversy over whether the role of data papers is to supplement or replace research papers [13]. For now, the value of data papers has been shown to be for promoting data sharing and reuse

Fig. 2. Results of the cluster analysis of interview themes on submission experiences and the usefulness of data related to data journals.

rather than for obtaining academic recognition, which is complementary to existing research papers. As suggested in previous studies, the publication of data papers is useful for ensuring data quality via peer review, facilitating follow-up research and reuse based on detailed technology provided in data papers, evaluating performance through the citability of data and data papers, and providing incentives for data-sharing [5,9]. Most of the interviewees agreed that the publication of data papers was useful in these terms. In other words, all of the interviewees agreed that the reliability, validity, accuracy, and reusability of data can be secured by the publication of data papers since the data in repositories linked to data papers are verified through the peer review process. The interviewees also noted that detailed descriptions about the data in data papers allow follow-up studies and that routine management and preservation of data is possible due to the availability of quality-controlled data from data papers. In addition, the researchers believed that a major function of data papers is that they helped to disseminate research achievements according to their citability. Furthermore, the economic value invested into original data production, collection, and analysis can be shared across disciplines and used as data sets, leading to further scientific progress through interdisciplinary research (Fig. 2).

Limitations: This was a qualitative study with a small number of subjects. Although the results may provide information on

researchers' general perceptions of data journals, additional quantitative research is required to obtain more generalizable results. Furthermore, the subjects were limited to senior-doctoral level researchers. A wider range of researchers could provide other opinions.

Conclusion: Data journals provide researchers with incentives for data-driven research. Data that have been validated through a peer review process can be used universally and preserved in a repository, and data that have been described in detail in data journals can be trusted and utilized in other studies. Data journals are expected to be used to promote subject expertise related to data, and pathways to facilitate interdisciplinary research based on data are expected to play an active role in accelerating scientific and technological advances and improving academic communication. However, we suggest that data journals should be developed as platforms for promptly providing verified, high-quality data needed by the scientific community, rather than for sharing performance-driven research achievements that are disclosed before the data are analyzed due to pressures related to research performance.

Conflict of Interest

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

Funding

This study was funded by the Korea Institute of Science and Technology Information (KISTI) (contract no. P19032).

References

1. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. *PLoS One* 2011;6:e21101. <https://doi.org/10.1371/journal.pone.0021101>
2. Sayogo DS, Pardo TA. Exploring the determinants of scientific data sharing: understanding the motivation to publish research data. *Gov Inf Q* 2013;30:S19-31. <https://doi.org/10.1016/j.giq.2012.06.011>
3. 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>
4. Edmunds SC, Li P, Hunter CI, et al. Experiences in integrated data and research object publishing using GigaDB. *Int J Digit Libr* 2017;18:99-111. <https://doi.org/10.1007/s00799-016-0174-6>
5. 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;6:2. <https://doi.org/10.2218/ijdc.v6i2.205>
6. Seo S, Kim J. Data journals: types of peer review, review criteria, and editorial committee members' positions. *Sci Ed* 2020;7:130-5. <https://doi.org/10.6087/kcse.207>
7. Kim J. An analysis of data paper templates and guidelines: types of contextual information described by data journals. *Sci Ed* 2020;7:16-23. <https://doi.org/10.6087/kcse.185>
8. Kim S, Chung E, Lee JY. Latest trends in innovative global scholarly journal publication and distribution platforms. *Sci Ed* 2018;5:100-12. <https://doi.org/10.6087/kcse.133>
9. 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>
10. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245-51. <https://doi.org/10.1097/ACM.0000000000000388>
11. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276-82. <https://doi.org/10.11613/BM.2012.031>
12. National Science Foundation. NSF's public access plan: today's data, tomorrow's discoveries—increasing access to the results of research [Internet]. National Science Foundation: Alexandria, VA; 2015 [cited 2021 Jul 17]. Available from: <https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>
13. Schopf J, Farace D, Prost H, Zane A. Data papers as a new form of knowledge organization in the field of research data. *KO Knowl Organ* 2019;46:622-38. <https://doi.org/10.5771/0943-7444-2019-8-622>

Appendix 1. Interview questionnaire

The questionnaire sought to elicit participants' opinions on motivations, needs, and barriers to publishing data papers and sharing data.

- (1) What are your motivations for publishing data papers? How are the submission and peer-review processes handled, and what obstacles do you encounter during these processes?
- (2) Why is it necessary to publish data papers? What are the differences between publishing data papers in data journals and traditional scholarly journals in terms of content, data quality, and the promotion of data reuse and academic development?
- (3) How significant do you think data papers will be in the future? Do you intend to publish more data papers? When considering publishing data papers, what are your major concerns (e.g., recognition of research results)?
- (4) What are the possibilities and limitations of founding a data journal in South Korea?
- (5) Please share your views of data sharing in terms of its needs, ways to access data (e.g., repositories), its scope, research funder requirements, and the barriers you face in sharing your own data.