



Recent trends in medical journals' data sharing policies and statements of data availability

Sun Huh

Ethics Editor, *Archives of Plastic Surgery*
Department of Parasitology and Institute of Medical Education, Hallym University College of Medicine, Chuncheon, Korea



Data sharing is defined as “making data available to people other than those who have generated them” [1]. It has become a mandatory process in scientific journals in a variety of fields, including oceanology, ecology, and genomics [2]. In those fields, specific data repositories are recommended as part of journals' policies. For example, sequence data of genes and proteins are usually deposited in the United States National Center for Biological Information (NCBI) database or other designated field-specific databases. Besides those specific types of data, sharing of data in general is also required. The adoption of data sharing policies may have been accelerated by the third version of the “Principles of Transparency and Best Practice in Scholarly Publishing” drafted by the Committee on Publication Ethics, the Directory of Open Access Journals, the Open Access Scholarly Publishers Association, and the World Association of Medical Editors, which includes “journal policies on data sharing and reproducibility” as the fourth sub-item under publication ethics. Only 227 (29.1%) of 781 academic society-published journals listed in the Science Citation Index Expanded fulfilled this sub-item [3]. This sub-item does not mean that adoption of data sharing is mandatory; instead, it refers to whether a journal's data sharing policy has been announced by its publisher or editor. Satisfying this sub-item is a prerequisite for application to PubMed Central, administered by the United States National Library of Medicine (NLM), which looks for ongoing publisher conformance with Principles of Transparency and Best Practice in Scholarly Publishing [4]. Therefore, even if an editor does not plan on adopting a data sharing policy, he or she should announce the “journal

policies on data sharing and reproducibility.”

This editorial aims to help editors prepare for adopting a data sharing policy and making an announcement on their journal's data sharing policy. Specifically, it addresses the following points: the advantages and limitations of data sharing policies, the present situation of data sharing policies in Korea, the various levels of data sharing, repository sites for general data sharing, descriptions of statements of data availability in the instructions to authors, in-text descriptions of datasets, and citations of data repositories in the References section.

ADVANTAGES AND LIMITATIONS OF DATA SHARING POLICIES

Several advantages/benefits of data sharing are well known. First, data sharing promotes reproducibility, which is an important issue regarding scientific articles. Even though data sharing cannot always guarantee full reproducibility of the results, it can at least prevent sloppy science and ensure that researchers uphold all relevant standards. Second, it helps to ensure scientific soundness. Of course, raw data themselves can be fabricated or falsified, which is a distinct type of research misconduct. Nonetheless, it is vital to minimize misconduct in scientific research in order to promote public confidence in science. Third, data sharing is beneficial for subsequent meta-analyses or mega-analyses. Analyzing the underlying data may be able to produce deeper analyses than is possible by analyzing the results of articles, especially for cohort studies using the same protocol or

randomized controlled studies. New concepts or conclusions can be produced from new analyses of data [5]. Fourth, data sharing can save resources by avoiding the need to regenerate the same data.

There is still scant evidence that data sharing promotes scientific development in clinical medicine. First, there are still only a few clinical studies that have used data sharing; second, ensuring a common data structure and consistency in the data has been a challenge; and third, no research design for sharing raw data has yet been established. In biomedical fields, it is mandatory to deposit DNA and protein sequence data in the NCBI database to publish articles. All data in the NCBI database is shared with all researchers in the world. However, it is still not mandatory to deposit clinical data. The International Committee of Medical Journal Editors (ICMJE) recently published Data Sharing Statements for Clinical Trials, which include the following requirement: “As of July 1, 2018 manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement.” This does not refer to mandatory data sharing, but instead means that authors must disclose the level of data sharing [6]. Therefore, it will be necessary to conduct further studies to determine the effects of data sharing in a variety of fields.

PRESENT SITUATION OF DATA SHARING POLICIES IN KOREA

It is not known how many scientific journals published in Korea have adopted a data sharing policy. The only available data on the present situation were obtained from a survey conducted in Korea from December 2018 to January 2019, and can be summarized as follows: “Of the 100 respondents (from 100 journals), 13 stated that their journals had already adopted a data sharing policy. Harvard Dataverse (2) and Mendeley Data (1) were chosen by the journals as a repository sites. The strength of the policy was recommendation-only in 10 of those 13 journals” [7]. Editors who adopted this policy did so because it is an international trend. Data sharing policies remain uncommon in Korea.

LEVEL OF DATA SHARING

The spectrum of data sharing can be classified as follows according to the level of data disclosure (Table 1) [8]. In a previous survey in Korea, only one journal had adopted mandated data sharing and peer-reviewed data [7,9].

PUBLIC REPOSITORIES FOR DATA TO BE DEPOSITED BY AUTHORS OR EDITORS/PUBLISHERS

Data can be deposited on a journal’s homepage after being submitted by the authors, or can be deposited into public repositories either by the authors themselves or by the editors/publishers. If an author deposits the data directly, the digital object identifier (DOI) of the data should be verified in the manuscript. If editors/publishers deposit authors’ data in a repository after the acceptance decision, the DOI is added by the repository. There are more than 2,000 repositories in various subject areas that can be searched through the Registry of Research Data Repositories (<https://www.re3data.org>) or the Repository Finder (<https://repositoryfinder.datacite.org/>). Internationally known repositories for depositing general data are presented in Table 2. There is no general data repository in Korea yet.

HOW TO DESCRIBE THE STATEMENT OF DATA AVAILABILITY IN THE INSTRUCTIONS TO AUTHORS?

An example from the *Journal of Educational Evaluation for Health Professions* (<https://www.jeehp.org/authors/authors.php>) is shown below:

Open data policy

For clarification on result accuracy and reproducibility of the results, raw data or analysis data will be deposited to a public repository, for example, Harvard Dataverse (<https://dataverse.harvard.edu/dataverse/jeehp/>) after acceptance of the manu-

Table 1. Various standardized data sharing policy categories [8]

Data sharing	Data availability statement is published	Data have been shared	Data have been peer-reviewed	Example
Encourages data sharing	Optional	Optional	Optional	-
Expects data sharing	Required	Optional	Optional	<i>British Journal of Social Psychology</i>
Mandates data sharing	Required	Required	Optional	<i>Ecology and Evolution</i>
Mandates data sharing and peer-reviews data	Required	Required	Required	<i>Geoscience Data Journal, American Journal of Political Science</i>

Table 2. List of generalist repositories

Repository name	Information on fees/costs	Size limits
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated
Figshare	Upload files up to 5 GB more than any other free offering for academic data publishing (20 GB of free private space to store research privately until the researcher chooses to make it public)	1 TB per dataset
Harvard Dataverse	Free of charge, donation Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded
Zenodo	Free of charge Donations towards sustainability encouraged	50 GB per dataset
Mendeley Data	Free of charge Contact repository for datasets over 10 GB	10 GB per dataset

Available from: <https://www.nature.com/sdata/policies/repositories#general>.

script. Therefore, submission of the raw data or analysis data is mandatory. If the data is already a public one, its URL site or sources should be disclosed. If data cannot be publicized, it can be negotiated with the editor. If there are any inquiries on depositing data, authors should contact the editorial office.

Clinical data sharing policy

This journal follows the data sharing policy described in “Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors” (<https://doi.org/10.3346/jkms.2017.32.7.1051>). Below is an example from *Archives of Plastic Surgery (APS)* (<https://www.e-aps.org/authors/authors.php>).

Data sharing

APS encourages data sharing wherever possible, unless this is prevented by ethical, privacy, or confidentiality matters. Authors wishing to do so may deposit their data in a publicly accessible repository and include a link to the DOI within the text of the manuscript.

Clinical Trials: APS accepts the ICMJE Recommendations for data sharing statement policy. Authors may refer to the editorial, “Data Sharing statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors,” in the *Journal of Korean Medical Science* (<https://doi.org/10.3346/jkms.2017.32.7.1051>).

HOW TO DESCRIBE DATA SHARING/AVAILABILITY IN AN ARTICLE?

Description of data sharing/availability varies according to the journal. Three examples are presented below:

(1) Data sharing on the journal homepage without deposition in a repository, in *PLoS One* [10]

Associated Data

Supplementary Materials

S1 Table: Detailed characteristics of retractions from Korean medical journals indexed in the KoreaMed Database. (XLSX)

pone.0163588.s001.xlsx (45 K)

GUID: 08F2D81E-658C-43AE-8E5D-35929A6DFF06

Data Availability Statement

All relevant data are within the paper and its Supporting Information files.

(2) Data sharing in a repository to which the authors deposited the data, in *Journal of the Korean Medical Association* [11]

Supplementary Materials

Supplementary materials are available from <https://doi.org/10.7910/DVN/X6CI4I>, Harvard Dataverse.

(3) Data sharing in a repository to which the editor/publisher deposited the data after acceptance of the manuscript, in *Journal of Educational Evaluation for Health Professions* [12]

Data availability

Data files are available from Harvard Dataverse: <https://doi.org/10.7910/DVN/T6WC1T>

Dataset 1. Dichotomous data converted from raw data of the items used in the 2nd cycle of evaluation and accreditation of medical schools by the Korea Institute of Medical Education and Evaluation from 2007 to 2011.

Dataset 2. Content of the items used in the 2nd cycle evaluation and accreditation.

Supplementary materials

Supplementary materials are available from Harvard Dataverse: <https://doi.org/10.7910/DVN/T6WC1T>

Supplement 1. Audio recording of the abstract.

In this journal, data were separately described from the supplementary materials. How best to distinguish between data and supplementary materials remains a challenge, as data can be included in the supplementary materials, and the converse is also possible.

In PubMed Central, data are listed in the section heading “Associated Data,” as in the following example [13]:

Associated Data

Supplementary Materials

Supplement 1. Data files are available from <https://doi.org/10.7910/DVN/WBTRMW>.

Supplement 2. Audio recording of the abstract.
jeehp-16-25-abstract-recording.avi (2.8 M)

GUID: 8FAC4837-D07D-4728-9594-1D677C2B6BED

Editors could adopt any of the above descriptions. Alternatively, a new format can be introduced by the editor. Each editor should choose an appropriate description of data sharing/availability. It suffices for the readers to be able to access the data, although the description can be formatted in a variety of ways.

HOW TO PRESENT CITED DATA IN THE REFERENCES SECTION

In the main text, any other data that are not included in the present article should be listed in the References section. Several different reference style formats exist, such as AMA (American Medical Association) style, APA (American Psychological Association) style, and NLM Citing Medicine. Below is the example of a reference according to NLM Citing Medicine, which is commonly used in biomedical journals [14]:

1. Lim S, Huh S. Goodness of fit of the items used in the 2nd cycle of evaluation and accreditation of medical schools by the Korea Institute of Medical Education and Evaluation based on the Rasch model [dataset]. 2019 [cited year month date]. In: Harvard Dataverse v.3 [Internet]. Cambridge, MA: Harvard College. Available from: <https://doi.org/10.7910/DVN/T6WC1T>.

Editors of Korean journals have recently begun to declare data sharing policies. Data sharing after publication is still not a mandatory provision in most biomedical journals in Korea that follow ICMJE policies. For editors, it is sufficient to understand how to announce their journals' data sharing policy and to indicate the level of sharing, how to deposit sharable materials into data repositories, and how to cite and describe data in their journals.

NOTES

Conflict of interest

SH has been an ethics editor since 2012. He was not involved in the peer reviewer selection, evaluation, or decision process of this editorial. No other potential conflicts of interest relevant to this editorial were reported.

ORCID

Sun Huh <https://orcid.org/0000-0002-8559-8640>

REFERENCES

1. U.S. Department of Energy. EERE digital data management glossary [Internet]. Washington, DC: Energy Efficiency & Renewable Energy [cited 2019 Oct 16]. Available from: <https://www.energy.gov/eere/funding/eere-digital-data-management-glossary>.
2. Kim J. Overview of disciplinary data sharing practices and promotion of open data in science. *Sci Ed* 2019;6:3-9.
3. Cho HW, Choi YJ, Kim S. Compliance of “Principles of transparency and best practice in scholarly publishing” in academic society published journals. *Sci Ed* 2019;6:112-21.
4. U.S. National Library of Medicine (NLM). Publisher practices [Internet]. Bethesda, MD: U.S. NLM [cited 2019 Jul 5]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#pubpract>.
5. Longo DL, Drazen JM. Data sharing. *N Engl J Med* 2016; 374:276-7.
6. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *J Korean Med Sci* 2017;32:1051-3.
7. Kim SY, Yi HJ, Huh S. Current and planned adoption of data sharing policies by editors of Korean scholarly journals. *Sci Ed* 2019;6:19-24.
8. John Wiley & Sons, Inc. Wiley's data sharing policies [Internet]. Hoboken, NJ: John Wiley & Sons, Inc. [cited 2019 Oct 16]. Available from: <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>.
9. Huh S. Establishment of an open data policy for Journal of Educational Evaluation for Health Professions, appreciation for invited reviewers, and acknowledgement of volunteers who made audio recordings. *J Educ Eval Health Prof* 2017; 14:37.
10. Huh S, Kim SY, Cho HM. Characteristics of retractions from Korean Medical Journals in the KoreaMed Database: a

- bibliometric analysis. PLoS One 2016;11:e0163588.
11. Vital Statistics Division, Statistics Korea, Shin HY, et al. Child injury death statistics from 2006 to 2016 in the Republic of Korea. J Korean Med Assoc 2019;62:283-92.
 12. Lim MS, Huh S. Goodness of fit of the items used in the 2nd cycle of evaluation and accreditation of medical schools by the Korea Institute of Medical Education and Evaluation based on the Rasch model. J Educ Eval Health Prof 2019; 16:28.
 13. Gelston CD, Patnaik JL. Ophthalmology training and competency levels in care of patients with ophthalmic complaints in United States internal medicine, emergency medicine and family medicine residents. J Educ Eval Health Prof 2019;16:25.
 14. Patrias K. Citing medicine: the NLM style guide for authors, editors, and publishers [Internet]. Bethesda, MD: U.S. National Library of Medicine; c2007 [cited 2019 Oct 16]. Available from: <https://www.nlm.nih.gov/citingmedicine>.
-
- Correspondence:** Sun Huh
Department of Parasitology and Institute of Medical Education, College of Medicine, Hallym University, 1 Hallimdaehak-gil, Chuncheon 24252, Korea
Tel: +82-33-248-2652, E-mail: shuh@hallym.ac.kr
Received: October 16, 2019 • Revised: October 23, 2019 • Accepted: October 24, 2019
pISSN: 2234-6163 • eISSN: 2234-6171
<https://doi.org/10.5999/aps.2019.01515> • Arch Plast Surg 2019;46:493-497
-