Data sharing attitudes and practices of researchers in Korean government research institutes: a survey-based descriptive study

Jihyun Kim¹, Hyekyong Hwang², Youngim Jung²,³, Sung-Nam Cho², Tae-Sul Seo²

¹Department of Library and Information Science, Ewha Womans University, Seoul; ²Korea Institute of Science and Technology Information (KISTI), Seoul; ³Department of HPC and Data Science, KISTI School, University of Science and Technology, Seoul, Korea

Abstract

Purpose: This study explored to what extent and how researchers in five Korean government research institutes that implement research data management practices share their research data and investigated the challenges they perceive regarding data sharing.

Methods: The study collected survey data from 224 respondents by posting a link to a SurveyMonkey questionnaire on the homepage of each of the five research institutes from June 15 to 29, 2022. Descriptive statistical analyses were conducted.

Results: Among 148 respondents with data sharing experience, the majority had shared some or most of their data. Restricted data sharing within a project was more common than sharing data with outside researchers on request or making data publicly available. Sharing data directly with researchers who asked was the most common method of data sharing, while sharing data via institutional repositories was the second most common method. The most frequently cited factors impeding data sharing included the time and effort required to organize data, concerns about copyright or ownership of data, lack of recognition and reward, and concerns about data containing sensitive information.

Conclusion: Researchers need ongoing training and support on making decisions about access to data, which are nuanced rather than binary. Research institutes’ commitment to developing and maintaining institutional data repositories is also important to facilitate data sharing. To address barriers to data sharing, it is necessary to implement research data management services that help reduce effort and mitigate concerns about legal issues. Possible incentives for researchers who share data should also continue to be explored.

Keywords

Information dissemination; Open research data; Restricted data sharing; Korean government research institutes; Barriers to data sharing
Introduction

Background/rationale
Many regulatory efforts to promote data sharing have been made and are underway at the international and national levels for the ultimate purpose of advancing science and research. High-level bodies and funding agencies are the primary contributors to the culture and practice of data sharing [1]. For example, the European Commission implemented the Open Research Data (ORD) pilot in Horizon 2020 (H2020), which requires H2020-funded projects to develop data management plans (DMPs) and provide open access to research data [2]. The principle of ORD is that research data should be “as open as possible, as closed as necessary,” and it emphasizes sound data management rather than forcing all research data to be open [3]. The US National Institute of Health (NIH) released a new Data Management and Sharing (DMS) policy in October 2020, effective as of January 25, 2023, which requires all applicants to submit DMS plans if the proposed research generates scientific data. Similar to the principle of ORD, the new NIH policy intends to “encourage data sharing to the extent that it is possible” [4]. In addition, publishers and individual journals influence researchers’ data sharing behavior by establishing data sharing policies that encourage or require making data available along with the publication of research articles [5,6]. Data sharing and reproduction is one of the policies on publication ethics that the recently updated version of the Principles of Transparency and Best Practice in Scholarly Publishing asks journals to include [7].

Compared to Europe and the United States’ commitment arising from the multiplicity of players in DMS, South Korea remains in the early phases of ORD agenda implementation. A national policy on research data management (RDM) is currently specified in the National Research & Development (R&D) Information Processing Standard. It is an administrative rule enforced by the ordinance of the Ministry of Science and ICT (MSIT) under the National R&D Innovation Act enacted in 2021. However, the policy on RDM is only applied to national R&D projects for which central government agencies consider it necessary to submit DMPs, and data sharing is not even mentioned in the policy [8]. Despite these limitations, government-funded research institutes under the National Research Council of Science and Technology (NST) affiliated with the MSIT have been directly affected by the policy and have gradually adopted DMPs; as of 2021, three of the 25 research institutes under the NST have implemented DMPs and data repositories. These include the Korea Institute of Geoscience and Mineral Resources (KIGAM), the Korea Institute of Oriental Medicine (KIOM), and the Korea Institute of Science and Technology Information (KISTI) [9]. In addition, the Korea Research Institute of Standards and Science (KRISS), another research institute under the NST, operates the National Standard Reference Data Center (NSRDC) [10]. Moreover, the National Institute of Ecology (NIE), a leading government research institute affiliated with the Ministry of Environment, has developed a platform for sharing ecological data named EcoBank [11].

Since the regulatory basis for RDM has been formed in South Korea and corresponding practices have been initiated in several government research institutes, it would be useful to investigate how researchers in the institutes regard data sharing. Such a study would lead to a better understanding of researchers’ experiences and what makes them reluctant when considering data sharing. A few studies have examined perceptions and the status of DMS based on surveys and/or interviews of researchers in Korean government research institutes [12–14]. However, the existing studies surveyed researchers in a single institute or were conducted before the regulations of RDM were established. Exploring the attitudes toward data sharing of researchers in multiple government research institutes where RDM practices are implemented will help address researchers’ needs regarding their institutes’ data management practices.

Objectives
This study examined the current data sharing practices and perceptions of researchers in five Korean government research institutes currently involved in RDM, which are KIGAM, KIOM, KISTI, KRISS, and NIE. The study assumed that researchers of these institutes have a certain level of understanding about data sharing and thus that it would be appropriate to recruit them as survey participants. The research questions were as follows: (1) To what extent do the researchers share their research data? (2) In what ways do they share research data? (3) What challenges do they perceive in relation to data sharing?

Methods

Ethics statement
This study was exempt from deliberation by the Institutional Review Board because there was no collection of sensitive information or individual identification information.

Study design
This is a descriptive study based on an online survey.

Data collection methods
A survey questionnaire was developed based on studies recently performed regarding data sharing and publication [15–17]. The questionnaire consisted of 29 questions in four...
Data sharing of researchers in Korean government research institutes

areas: (1) research data creation and management; (2) data sharing and publication; (3) perceptions of data publication; and (4) demographic information. SurveyMonkey was used to construct an online questionnaire, and a link was distributed via a discussion forum on the homepage of each of the five government research institutes from June 15 to 29, 2022. As a result, 224 responses were collected and used for the analysis. Almost all responses were complete, except for eight responses that failed to provide demographic information.

Units of study
The unit of study in this research was the individual, since the survey data were collected from 224 researchers employed in the five Korean government research institutes.

Data analysis
This study focused on analyzing responses concerning the extent, methods, and barriers to data sharing. Descriptive statistical analyses were performed to compare the survey responses in terms of whether and to what extent respondents shared research data, which of the various ways of sharing data they utilized, and the challenges they perceived.

Results
Participants
The demographics of the respondents are presented in Table 1. The majority of respondents were men and had doctoral degrees. Most were in their 30s or 40s. Regarding the disciplines of the respondents, most were in 56 information communications technology (25.9%), 38 biological science (17.6%), 21 public health and medicine (9.7%), and 16 earth science (7.4%). In total, 24 disciplines were reported by the respondents (Fig. 1).

The extent of data sharing
Out of 224 respondents, 180 (80.4%) answered that they had experience collecting or creating research data. Among those 180 researchers, 32 mentioned that they had never shared their research data; thus, the remaining 148 respondents had shared at least some of their data (Table 2). Only seven respondents stated that they shared all of their data, while the majority shared some or most of theirs.

The 148 respondents with data sharing experience were also asked with whom they had shared research data. As shown in Table 3, sharing data with principal investigators or coinvestigators involved in creating or collecting data received the greatest proportion of the responses (n = 94, 36.3%), followed by sharing data with all participants in research projects where data were collected/created (n = 86, 33.2%). The findings indicate that restricted data sharing—“the exchange of data between the members of a specific group or project consortium to their mutual benefit” [2]—was more common than data sharing on request (n = 60, 23.2%) or making data publicly accessible (n = 19, 7.3%).

This study also analyzed the responses presented in Table 3 to compare the proportion of respondents who only shared data in a restricted way with those who shared data on request or created an ORD. The responses were collected from a question that allowed multiple answers. Thus, the study identified the number of respondents providing one or two answers who selected only the choices categorized under restricted data sharing. This response type was found to constitute a majority, with a total of 80 such respondents (54.1%) (Table 4). This result indicates that although 148 respondents reported sharing their data, most conducted data exchange with a limited scope (i.e., within a research project) rather than making data available to outside researchers who needed it or to anyone at any time.

Methods of data sharing
In terms of the ways in which the respondents reported sharing their data, 110 responses (44.0%) described sharing data directly to individual researchers on request. Strongly recommended data sharing methods, such as deposition within a data repository, were less likely to be used by the respondents. However, sharing data through an institutional repository or server received the second greatest number of responses (n = 45, 18.0%). This might have resulted from the fact that the research
institutes with which these respondents are affiliated have implemented data repositories and DMPs. A personal or research team website was the third most used method of sharing data (n = 40, 16.0%). Sharing data via journals’ databases or repositories while publishing research articles also received 37 responses (14.8%). This result highlights the role of journals in facilitating data sharing. Only a small portion of responses indicated sharing data via disciplinary repositories or data journals. These results are illustrated in Table 5.

### Barriers to data sharing

One survey question asked about factors that made respondents reluctant to share data. This question was asked of the 180 respondents with experience creating or collecting research data, and multiple answers were permitted. As shown in Table 6, the top three reasons received a similar proportion of responses (around 16%). These reasons were the time and effort required to organize data, unclear copyright or ownership of data, and a lack of reward or recognition for sharing data. The factor of concerns about data having sensitive information was
also selected in 77 responses (15.3%), followed by concerns about data having errors (n = 60, 11.9%). Other options selected by participants included no regulations for data sharing (n = 43, 8.5%), the lack of an adequate data sharing platform (n = 35, 7.0%), the perceived lack of need for their data (n = 23, 4.6%), and the absence of funding for sharing data (n = 17, 3.4%) (Table 6).

### Discussion

#### Key results
Among the 148 respondents who had experienced data sharing, the vast majority reported sharing some or most of their data. Sharing data with principal investigators, coinvestigators, or participants in a research project was more common than sharing data with outside researchers on request or creating an ORD. A majority of the 148 respondents with data sharing experience shared data only in a restricted manner. Therefore, the extent of data sharing performed by the respondents is limited in light of international efforts to promote ORD.

The most common method of sharing data was to provide data directly to individual researchers who asked for it, even though funding agencies and journals strongly recommend depositing data in repositories. Yet, institutional repositories or servers were used the second most. This might be because the research institutes with which the respondents were affiliated operated institutional repositories as data sharing platforms. Journals’ databases or repositories were also employed; thus, journals’ role in promoting data sharing should be recognized.

Various factors that impeded data sharing were identified. The most commonly indicated factor was the time and effort required to organize data, followed by concerns about copyright or ownership of data and lack of reward or recognition for data sharing. Respondents also had concerns about sharing data containing sensitive information and the possibility of errors in their data. Exploring ways to mitigate such barriers to data sharing is necessary at an institutional and national level.

#### Interpretation
Most respondents with data sharing experience made some or most of their data available to others. This result is similar to that of the study by Tenopir et al. [16], which conducted an international survey of researchers’ DMS practices. However, the proportion of respondents in this study who reported sharing all of their data was about four times lower than that of Tenopir et al. [16]. Restricted data sharing was also more common than making data available to outside researchers or everyone. According to the principles of the H2020 ORD pilot and newly released NIH policies, ORD is strongly recommended, but restricted data sharing is reasonable if more value can be provided by restricting data access [2]. Since decisions about

---

**Table 4.** Proportion of respondents who participate in restricted data sharing only versus in other types

<table>
<thead>
<tr>
<th>No. of responses</th>
<th>Restricted data sharing only</th>
<th>Data sharing on-request or ORD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48 (60.0)</td>
<td>16 (23.5)</td>
<td>64 (43.2)</td>
</tr>
<tr>
<td>2</td>
<td>32 (40.0)</td>
<td>29 (42.6)</td>
<td>61 (41.2)</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>19 (27.9)</td>
<td>19 (12.8)</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>4 (5.9)</td>
<td>4 (2.7)</td>
</tr>
<tr>
<td>Total</td>
<td>80 (54.1)</td>
<td>68 (45.9)</td>
<td>148 (100)</td>
</tr>
</tbody>
</table>

Values are presented as number (%). ORD, Open Research Data.

**Table 5.** How respondents share data

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share data directly at the request of an acquaintance or individual researcher</td>
<td>110 (44.0)</td>
</tr>
<tr>
<td>Through an institutional repository or server</td>
<td>45 (18.0)</td>
</tr>
<tr>
<td>Through a personal or research team website</td>
<td>40 (16.0)</td>
</tr>
<tr>
<td>Through a database or repository of a journal when publishing a research article</td>
<td>37 (14.8)</td>
</tr>
<tr>
<td>Through a data archive or repository in my discipline</td>
<td>11 (4.4)</td>
</tr>
<tr>
<td>Through a database or repository of a data journal when publishing a data paper</td>
<td>7 (2.8)</td>
</tr>
<tr>
<td>Total</td>
<td>250 (100)</td>
</tr>
</tbody>
</table>

*Multiple answers allowed.*

**Table 6.** What makes respondents reluctant to share data

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time and effort required to organize data</td>
<td>84 (16.7)</td>
</tr>
<tr>
<td>Copyright or ownership of data is unclear</td>
<td>83 (16.5)</td>
</tr>
<tr>
<td>Lack of adequate reward or recognition for sharing data</td>
<td>81 (16.1)</td>
</tr>
<tr>
<td>Data contains sensitive information (e.g., personal information)</td>
<td>77 (15.3)</td>
</tr>
<tr>
<td>Data may contain errors</td>
<td>60 (11.9)</td>
</tr>
<tr>
<td>No data sharing obligations or related regulations</td>
<td>43 (8.5)</td>
</tr>
<tr>
<td>Unable to find the right platform to share my data</td>
<td>35 (7.0)</td>
</tr>
<tr>
<td>Perceived lack of need for data</td>
<td>23 (4.6)</td>
</tr>
<tr>
<td>No funding for sharing data</td>
<td>17 (3.4)</td>
</tr>
<tr>
<td>Total</td>
<td>503 (100)</td>
</tr>
</tbody>
</table>

*Multiple answers allowed.*
access to data are nuanced rather than binary [2], it is important for Korean government research institutes to provide ongoing training and support for affiliated researchers so that they can make informed decisions and adequate justifications if restricted data sharing is needed. In addition, developing an appropriate infrastructure that enables researchers to share and preserve their data is necessary, and funding agencies and journals strongly recommend depositing data into robust repositories [4,5]. Therefore, research institutes should make commitments to implement data repositories and encourage the submission of DMPs to build good DMS practices.

Moreover, a recent systematic review of studies on factors associated with data sharing suggested 11 categories of factors: researcher’s background, requirements and formal obligations, intrinsic motivations, facilitating conditions, trust, expected performance, social influence and affiliation, effort, researcher’s experience, legislation and regulation, and data characteristics [18]. The findings of this study are mostly related to effort (time and effort required to organize data), legislation and regulation (concerns about copyright or ownership and sensitive information), and expected performance (lack of reward and recognition). To alleviate these factors, it is desirable to develop RDM services for researchers, which will reduce time and save effort in data preparation and organization. Furthermore, training and consultation services regarding copyright and privacy will help researchers better understand legal issues and be less concerned about accidentally violating the law when considering data sharing. Finally, as the lack of incentives has been suggested as a major impediment to data sharing [19], continuing efforts and discussions regarding sufficient rewards for data sharing are required within research institutes and externally in disciplines and governments.

Limitations
This study is based on descriptive statistical analyses only. As such, the results show the current state of data sharing behavior and perception reported by respondents, but an inferential interpretation is not possible.

Conclusion
This study found that survey respondents affiliated with Korean government research institutes commonly performed restricted data sharing. ORD is internationally recommended, but at the same time, restricted data sharing is allowable if it is possible to derive value while maintaining restricted access to data. Providing ongoing training and support for researchers can help them make informed decisions about access to data. It is also important for research institutes to develop and sustain institutional data repositories as a platform for DMS. To address barriers to data sharing, it is necessary to implement RDM services that help reduce the effort required to organize data and provide consultations for copyright and privacy issues. It is also necessary to continue exploring possible incentives for researchers who share data from institutional, disciplinary, and governmental perspectives.

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

Funding
This work was supported by the National Research Council of Science and Technology (No. NST-02).

References
1. Popkin G. Data sharing and how it can benefit your scientific career. Nature 2019;569;445–7. https://doi.org/10.1038/d41586-019-01506-x


15. Korea Social Science Data Archive (KOSSDA). A survey of researchers about research data creation and management. KOSSDA; 2021.


