
Research Data Collection Guidelines

2023. 5.

Table of Contents

1. Overview	1
2. Research Data Definition	2
3. Research Data Collection	5
4. Research Data Deposit	8
5. DMP Writing	10



List of Tables

〈Table 1〉 Dublin Core metadata elements	6
〈Table 2〉 Example of DMP writing (1)	10
〈Table 3〉 Example of DMP writing (2)	12



List of Figures

〈Figure 1〉 Data registration procedure when depositing research data 9



1

Overview

【 Purpose 】

- ☐ The presentation of research data collection guidelines applied by the Geoscience Data Center of the Korea Institute of Geoscience and Mineral Resources (KIGAM) when collecting research data.

【 Target 】

- ☐ Researchers and users who want to deposit research data to the Geo Big Data Open Platform.

【 Scope of Application 】

- ☐ Applies to research data during in-house research activities and research data donated by external organizations and individuals.

【 Application 】

- ☐ For matters not specified in these guidelines may be subject to the research data management guidelines of the National Research Council of Science & Technology (NST) and KIGAM Data Management Regulations.

【 Research Data Definition 】 ¹⁾

- ☐ “Research data” refers to factual data generated through various experiments, observations, investigations, and analyses of research projects, which are essential for the verification of research results.

【 Research Data Classification 】 ²⁾

- ☐ Research data is categorized into¹ primary data, secondary data, digitally converted data, metadata, external data, and collections.
- ☐ This excludes research notes, drafts of scientific papers, future research plans, communications with colleagues, and real physical materials, but includes metadata about real physical materials such as drill cores, rocks, and fossils in research data.
 - Primary data: This refers to data generated through various surveys, experiments, monitoring, observations, measurements, analyses, etc. of the research project, which are essential and objective factual data for reproducing research results.
 - Secondary data: This refers to data produced through processing and analysis based on primary data, such as tables, images, graphs, models, various geological maps, geophysical anomaly maps, and geochemical maps.
 - Digitally converted data: data that has been converted from analog to digital form.
 - Metadata: This refers to data used to describe the data, consisting of the title of the data required for the data description, the producer of the data, the equipment and method of production of the data, the content of the data, the location (location coordinates) and time of acquisition, the format of the data, the quality of the data, etc.
 - External data: This refers to data provided by external sources, both paid and unpaid.
 - Collection: Any aggregation of physical or digital resources.

1) Ministry of Science and ICT (2020). Article 2, Paragraph 6 of the National R&D Information Processing Standards.

<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=108&mPid=103&bbsSeqNo=83&nttSeqNo=3175298>

2) STAR Library. Research Data Management. <https://starlibrary.org/research-data-management>

【 Classification by Data Production Method Criteria 】

- ☐ Observational data: These are data produced through observation or monitoring.
- ☐ Experimental data: These are data produced through experimentation.
- ☐ Simulation data: These are data produced through simulation.
- ☐ Derived or compiled data: These are data produced through compilation or extraction.

【 Importance and necessity of sharing research data 】³⁾

- ☐ With the shift towards data-intensive research, which attempts to make scientific discoveries through vast amounts of research data produced by various measurement and experimental equipment, the importance of recycling and managing research data is increasing.
- ☐ With the open science and open access movements, data and publications generated by publicly funded research projects are shifting towards being published in public data repositories.
- ☐ Sharing research data increases its availability to other researchers or institutions.
- ☐ This is an opportunity to increase the reputation of individual researchers or research institutions through data citation, etc., and to develop better research through data verification, etc.
- ☐ Sharing research data can reduce the cost of duplicate production and publication of data, allowing for greater focus on future research.
- ☐ This allows for the sharing of most pre-processed research data while respecting privacy, research ethics, and other sensitive issues.
- ☐ If the research project management organization promotes the disclosure of

3) National Research Council of Science & Technology. Guidelines for Management and Utilization of Research Data.

<https://www.nst.re.kr/www/selectBbsNttView.do?SESSIONID=0C1DA6F31EBF881198E95FF8105B74AB?key=64&bbsNo=34&nttNo=2205&searchCtgry=&searchCnd=all&searchKwd=&integrDeptCode=&pageIndex=5>

research data through the Data Management Plan (DMP), standardization of data acquisition, storage, and metadata creation is the best way to manage research data to realize research data sharing.



【 Research data collection scope 】

- ☐ KIGAM collects and manages geological resource datasets for research fields such as national geology, mineral resources, oil seabed, and geological environment. This includes research data generated during research activities at the institute and data obtained from outside.
- ☐ The scope of research data collection (research data acquired through research activities) specified above includes DMP.
 - According to the Research Data Management Regulations of KIGAM, DMP must be submitted when conducting research projects, so it falls under the collection object that is managed and preserved as a research asset.
- ☐ The types of research data specified in the DMP are as follows:
 - Field investigation data
 - Outdoor exploration/measurement data
 - Sample and assay data
 - Map data (Final deliverables in the form of drawings, such as geologic maps and geologic thematic maps)
 - Course materials (International School for Geoscience Resources)
 - Literature/survey data
 - External data
 - Other-categorized data (data not included in the above categories)

【 Select and Evaluate Research Data 】

- ☐ Research data collection criteria:
 - Data with demonstrated tangible value to research and education and ongoing archival significance.
 - Data produced in core areas defined by the organization.
 - Data produced by agency-funded geology and resources projects.
- ☐ Security, privacy, and confidentiality considerations:
 - Otherwise, follow the Research Data Ethics • Copyright • Licensing Guidelines.
 - For sensitive data collected and stored, meet recognized standards for privacy and confidentiality.

- Licenses applicable to specific data collections should be managed in accordance with the law, given the repository's resources, goals, and mission.
- ☐ Copyright and licenses:
- Other details follow research data Ethics • Copyright • Licensing guidelines.
 - When collecting research data, the owner of intellectual property rights must be identifiable.
 - Individuals or organizations with intellectual property rights to submit research data must agree to the deposit conditions set by the repository.
 - Copyright holders of research data select and apply the appropriate license type when licensing data use.
 - CC licenses are prioritized for research data sharing and utilization.
- ☐ Research data quality:
- Encourage the collection of research data with comprehensive technical documentation that provides users with information to assess data quality and reliability when collecting research data.
 - Prefer research data in its original form.
- ☐ Metadata
- Metadata used during ingestion and deposition uses Dublin Core's format as shown in <Table 1>⁴⁾.

<Table 1> Dublin Core metadata elements

Element	Contents
Title	<ul style="list-style-type: none"> • A name given to the resource.
Creator	<ul style="list-style-type: none"> • An entity responsible for making the resource.
Type	<ul style="list-style-type: none"> • The nature or genre of the resource.
Contributor	<ul style="list-style-type: none"> • An entity responsible for making contributions to the resource.
Publisher	<ul style="list-style-type: none"> • An entity responsible for making the resource available.
Date	<ul style="list-style-type: none"> • A point or period of time associated with an event in the lifecycle of the resource.
Language	<ul style="list-style-type: none"> • A language of the resource.
Format	<ul style="list-style-type: none"> • The file format, physical medium, or dimensions of the

4) DCMI (1999) 'Dublin Core Metadata Element Set, Version 1.1'

National Library of Korea. DCMES Features. <https://librarian.nl.go.kr/LI/contents/L1020200000.do>

Element	Contents
	resource.
Description	<ul style="list-style-type: none"> • An account of the resource.
Subject	<ul style="list-style-type: none"> • A topic of the resource.
Relation	<ul style="list-style-type: none"> • A related resource.
Identifier	<ul style="list-style-type: none"> • An unambiguous reference to the resource within a given context.
Rights	<ul style="list-style-type: none"> • Information about rights held in and over the resource.
Source	<ul style="list-style-type: none"> • A related resource from which the described resource is derived.
Coverage	<ul style="list-style-type: none"> • The spatial or temporal topic of the resource, spatial applicability of the resource, or jurisdiction under which the resource is relevant.

【 Preferred Research Data Format 】

- ☐ Data that is available in a variety of computing and technology environments.
- ☐ Data that is available in an easily accessible format for users.
- ☐ Data that is accessible and usable without compromising research value.
- ☐ Data that can be converted to a format that can be used by various statistical or analytical software.
- ☐ Data that does not require to be interpreted using additional software.

【 Research Data Ethics 】

- ☐ Follow the Research Data Ethics • Copyright • Licensing Guidelines for other information.
- ☐ To prevent ethical issues that may arise in the production and collection of research data, KIGAM collects relevant data in compliance with research data ethics.

【 Meaning of Research Data Deposit 】

- ☐ Registering the research data produced in the system (repository) for the purpose of facilitating future reuse and continuous access to the research data.

【 Requisites for Research Data Deposit 】⁵⁾

- ☐ Ensuring continuous access to research data.
- ☐ Providing professional research data management, preservation, and access.
- ☐ Reducing storage costs and enabling stable storage of large amounts of data.
- ☐ Long-term management and preservation of research data in a secure environment.
- ☐ Can be utilized as a resource with potential value that can be reused in the future.

【 Materials to be submitted by researchers when depositing research data 】

- ☐ Research data files
 - The data should be in a state that allows software to process (verify, modify, convert, extract, etc.) the individual contents or internal structure of the research data.
 - It is recommended that the Research Dataset or Research Data File are provided in a universal format for future reuse, or in a specific format that is commonly used by the domain-specific community.
- ☐ Documentation
 - Provide documentation files with the study data files that are necessary to interpret the study data files.
 - Documentation files may include codebooks, data collection tools, summary statistics, project summaries, and a list of data-related publications.

5) National Research Council of Science & Technology (NST). Research on how to manage and utilize research data from government-funded research institutions.
<https://www.nst.re.kr/www/selectBbsNttView.do;JSESSIONID=0C1DA6F31EBF881198E95FF8105B74AB?key=64&bbsNo=34&nttNo=2205&searchCtgry=&searchCnd=all&searchKrwd=&integrDeptCode=&pageIndex=5>

☐ Metadata

- Metadata describing the contents of the research data file must be provided along with the research data file.
- The metadata format (DC) you use may include: title, creator, type, contributor, format, description, subject, and terms. (see Research Data Collection Guidelines <Table 1>).

【 Research Data Registration Procedure 】

☐ Registration

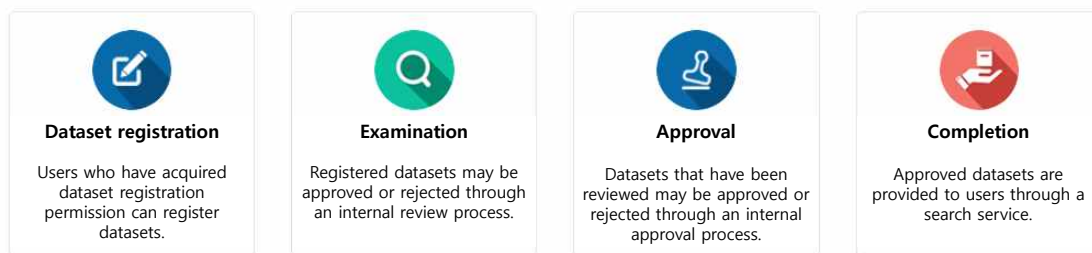
- Researchers check research data for quality and anomalies and submit research data to the data repository <Figure 1>.

☐ Approval and review

- Principal investigators or institutional/departmental research data management personnel review and approve submitted research data <Figure 1>.
- Key review points: Data format, compliance with metadata technical formats, compliance with research data disclosure and licensing, and inclusion of sensitive information.

☐ What to look for when making a deposit

- Checklist
 - Follow naming conventions
 - Check for typos
 - File size (maximum single file size limit: 100 GB)
 - Check the DMP taxonomy code



<Figure 1> Data registration procedure when depositing research data

5

DMP Writing

【 The Meaning of DMP 】

- ☐ DMP is a plan for the production, preservation, management, and joint utilization of research data.⁶⁾
- ☐ It means a document that is prepared and submitted together with the research plan when establishing a research plan for the collection, management, preservation, opening, and utilization of research data.

【 Purpose of DMP Writing 】

- ☐ To plan in advance the types of data to be produced in the research process and how to acquire them and to manage the data produced efficiently.

【 Create and Fulfill, Change 】

- ☐ Matters regarding the preparation and implementation of the DMP are specified in Article 13 (Preparation and implementation of the data management plan) of the Research Data Management Regulations of the KIGAM and include the subjects of preparation, and matters to be reviewed after preparation.
- ☐ Changes to the DMP are specified in Article 14 (Changes to the Data Management Plan) of the KIGAM's Research Data Management Regulations.

【 Configuration 】

- ☐ DMP can be written as shown in <Table 2>.

<Table 2> Example of DMP writing (1)

Project Title			
Research Period		Research Project Manager	
Open Status	Open/Close	Reasons	If closed, please write reasons
1. Field Survey Data			

6) Regulations On the Management Of National Research and Development Projects, Article 2 (Definitions)

ID	Classification Code	Title	Survey Method	Producer	
1-1					
1-2					
2. Field exploration/measurement data					
ID	Classification Code	Title	Exploration Method	Producer	
2-1					
2-2					
3. Sample and sample analysis data					
ID	Classification Code	Sample Name	Method of Analysis	Number of Sample	Producer
3-1					
3-2					
4. Geological thematic maps					
ID	Classification Code	Title	Scale	Producer	
4-1					
4-2					
5. Lecture materials					
ID	Classification Code	Title of Education Course	Title	Producer	
5-1					
5-2					
6. Literature/survey data					
ID	Classification Code	Title	Purpose of Survey	Producer	
6-1					
6-2					
7. External data					
ID	Classification Code	Title	Organization	Producer	
7-1					
7-2					
8. Other - unclassified data					
ID	Classification	Title	Contents	Producer	

	Code			
8-1				

☐ Metadata standards can be written as shown in <Table 3>.

<Table 3> Example of DMP writing (2)

Division	Content and examples
Metadata standards	<ul style="list-style-type: none"> • TTAS.IS-19115: Metadata Standard for Geographic Information Management • Dublin Core: A set of metadata elements standardized by ISO for efficient search and management of various digital resources on the Internet. • Darwin Core: An extension of Dublin Core for biodiversity information • ABCD(Access to Biological Collection Data): Standards for access and exchange of primary biodiversity data, including specimens and observations • ABCDEFG (Access to Biological Collection Databases Extended for Geosciences): An extension of ABCD for geoscience data

【 Expectations when DMP writing 】

- ☐ Easily track the creation, management, and sharing of research data generated through assignments.
- ☐ Ensure reproducibility of research data.
- ☐ Ensure research integrity by allowing the reuse of research data.
- ☐ Identify duplicate studies to improve the efficiency of a researcher' s study.

Version No.	Date	Contents
0.1	2023. 03. 20.	Create document outline
0.6	2023. 04. 28.	Create draft
0.8	2023. 05. 08.	Guideline review
1.0	2023. 05. 19.	Accept review comments

