
Research Data Management Guidelines

2023. 5.

Table of Contents

1. Overview	1
2. Research Data Management System	2
3. Research Data Collection	8
4. Research Data Storage	9
5. Research Data Quality Assurance	14
6. Research Data Description	22
7. Research Data Sharing and Reuse	24

List of Tables

〈Table 1〉 Data Lifecycle Component Details	5
〈Table 2〉 Deposited data preference format	9
〈Table 3〉 Version numbering scheme examples	12
〈Table 4〉 Findable Checklist	17
〈Table 5〉 Accessible Checklist	19
〈Table 6〉 Interoperable Checklist	20
〈Table 7〉 Reusable Checklist	21

List of Figures

〈Figure 1〉 Data Lifecycle for Research Data Management	5
〈Figure 2〉 Research Data Standard Citation Format	29

1

Overview

【 Purpose 】

- ☐ Presents data management guidelines for collection and disclosure of research data at the Geological Resources Data Center of the Korea Institute of Geoscience and Mineral Resources (KIGAM)

【 Target 】

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

【 Scope of Application 】

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

【 Application 】

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

2

Research Data Management System

【 Meaning of Research Data Management 】

- ☐ It refers to the methods of storing, sharing, accessing, preserving, and reusing data generated and collected during the conduction of a research project.
- ☐ It also includes Data Management Plans (DMPs), processes, and long-term storage and sharing.
- ☐ The definition of research data follows the contents of the KIGAM Data Collection Guide.
- ☐ All research data produced and retained by KIGAM's national research and development projects are managed, and any data subject to disclosure are managed in consideration of KIGAM's security and research data regulations.

【 Purpose and Goals 】

- ☐ The purpose of research data management is to promote the sharing of research data. This is achieved by ensuring long-term access and sustainable utilization of data and by enabling the reuse of research data in the future.
- ☐ To facilitate the reuse of research data, it is important to create data that can be understood and used appropriately, even by people not directly involved in the project.
 - This can be accomplished by creating a DMP and sharing basic data descriptions, along with the details of how data are managed and reused.
- ☐ Research data management helps
 - reduce the risk of research data loss;
 - improve efficiency and ease of research data control;
 - increase citations and future collaborations through improved research data visibility;
 - demonstrate research integrity and validate research results;
 - increase research impact through knowledge transfer;
 - advance research through the reuse of research data.

【 Principles 】

- ☐ Data should adhere to FAIR (Findable, Accessible, Interoperable, Reusable) principles.
 - Findable: Data should be assigned a globally unique and persistent identifier, and are described with rich metadata. Metadata clearly and explicitly includes the identifier of the data they describe. (Meta)data are registered or indexed in a searchable resource
 - Accessible: (Meta)data are retrievable by their identifier using a standardized communications protocol. The protocol is open, free, and universally implementable. The protocol allows for an authentication and authorization procedure, where necessary. Metadata is accessible, even when the data are no longer available.
 - Interoperable: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. (Meta)data use vocabularies that follow FAIR principles. (Meta)data include qualified references to other (meta)data
 - Reusable: (Meta)data are richly described with various accurate and relevant attributes. (Meta)data are released with a clear and accessible data usage license. (Meta)data are associated with detailed provenance. (Meta)data meet domain-relevant community standards.

【 Establishing Data Management System 】

- ☐ Policies and procedures
 - KIGAM defines research data management standards to help researchers conduct responsible research and manage data by adhering to the KIGAM Data Management Rules, KIGAM also establishes policies and procedures necessary for data management for the organization.
- ☐ Dedicated organization and manpower
 - KIGAM recruits professional personnel for the management and utilization of research data and establishes and operates a dedicated organization, the Research Data Department, which is specified in the KIGAM Data Management Rules (Chapter 3, Research Data Department).
 - In order to oversee and coordinate research data management, the KIGAM operates a research data department that performs all tasks related to the collection, management, preservation, opening and utilization of research data.
 - The main duties are as follows:

- Establishment of basic policies on research data management and improvement of the system.
 - Establishment and operation of research data management system.
 - Development, research, and dissemination of research data standardization, quality control, preservation, openness, and utilization technologies.
 - Operation of the Research Data Management Committee.
 - Providing education, training, counseling, and technical support on research data management.
 - Linkage, interact, and cooperation with other organizations for research data management.
 - Other matters necessary for research data management.
- KIGAM operates a research data management committee to deliberate on matters related to the DMP and research data, as specified in the KIGAM Data Management Rules (Chapter 2, Research Data Management Committee).
 - The committee is composed of up to 25 members, including the heads of each department, and the chairperson is the vice president. For efficiency of DMP deliberation and data management, the members shall be the same as the members of the Research Projects Review Committee.
- ☐ Definition of requirements
- For research data management, the size, type, production cycle and search requirements of research data produced or held by the KIGAM are defined.
 - For systematic management and wide sharing and utilization, we are preparing standardization measures for research data and processes.
- ☐ Research Data Infrastructure
- KIGAM provides IT infrastructure that is essential for stable research data management and long-term preservation, and for supporting the sharing and utilization of various research data.
- ☐ Support services
- KIGAM provides education, consultation, and support services to researchers for systematic research data management and utilization.

【 Data Lifecycle Management Methodology 】

- The research data management method in this guide follows the Data Lifecycle, which is shown in <Figure 1>¹⁾



<Figure 1> Data Lifecycle for Research Data Management

- <Table 1> describes the details of the Data Lifecycle components

<Table 1> Data Lifecycle Component Details

Phase	Component	Details
Collect	DMP writing	<ul style="list-style-type: none"> • Determine what data needs to be created or collected • Identify standards for the format and quality of data and metadata • Roles and responsibilities of data manager • Develop data sharing agreements • Understand data access controls and copyright • Explore and acquire existing data • Collect data (experiments, observations, measurements, simulations) • Metadata and documentation descriptions for the acquired data

1) Kim, J., Kim, S., & Jeon, Y. (2019). Data Life Cycle Proposal for Research Data Management. Journal of the Korean Society for Library and Information Science, 53(4), 309-340.

Phase	Component	Details
Store		<ul style="list-style-type: none"> • Store data in compliance with applicable standards • Create a short-term retention plan to minimize potential losses • Identify the number of copies of research data to be stored and how they will be synchronized • Provide storage for research data • Maintain a copy site for backup of data transfer systems stored in cloud-based services • Provide a data download service from a backup site in the event of a power outage • Provide criteria for comparing storage (storage space) solutions • Ensure integrity and accessibility when backing up data
Manage	Quality assurance	<ul style="list-style-type: none"> • Ensure authenticity, reliability, and availability while maintaining the integrity of the data • Ensure acceptable data structures or file formats • Quality assurance and control • Documenting of data quality • Data backup
	Security	<ul style="list-style-type: none"> • Principles for limiting disclosure risk • Data collection methods with limited use • Access restrictions (to protect privacy and confidentiality) • Data privacy anonymization • Protection of data from accidental loss, damage, and unauthorized access
Access & Use		<ul style="list-style-type: none"> • Ensure that both designated users and reusers have daily access to data • Apply access control and authentication procedures • Provide preferred file formats • Register projects on project sites and store data in open repositories

Phase	Component	Details
		<ul style="list-style-type: none"> • Support methods of discovery, analysis, repurposing, distribution, and presentation
	Preserve	<ul style="list-style-type: none"> • Implement measures to preserve and retain the administrative characteristics of your data for the long term • Develop a plan for long-term data retention • Determine what data to retain, where to retain it, and what documentation is needed with the data • Create metadata and documentation for preservation • Organizing and storing data
Share & Reuse	Citation	<ul style="list-style-type: none"> • Dataset and project documentation techniques • Document data history • Document data by explaining the its purpose (why), source (who), time of creation (when), location (where), and method of creation (how)
	Publication	<ul style="list-style-type: none"> • Catalog and describe • Ensure consistent identification of published data • Data release process • DOI, IGSN²⁾ ((International Generic Sample Number) issuance • Distribution and data sharing • Copyright settings • Dissemination of information about your data, promote your data

2) IGSN. <https://www.igsn.org/>

【 Data Collecting 】

- ☐ KIGAM has developed and complies with research data collection guidelines. The details of research data collection are described in the guidelines.
 - KIGAM Data Collection Guide includes the definition of research data, scope of collection, selection and evaluation, deposition, and organization and examples of DMPs.

【 Guide to Writing a DMP 】

- ☐ In the data collection phase of research, the following activities are required in addition to creating a DMP:
 - Determine what data should be created or collected.
 - Identify standards for the format and quality of data and metadata.
 - Roles and responsibilities of data manager.
 - Develop data sharing agreements.
 - Understand data access controls and copyright.
 - Explore and acquire existing data.
 - Data collection (experiments, observations, measurements, simulations).
 - Metadata and documentation techniques to the corresponding data.
- ☐ KIGAM collects and manages geoscience datasets such as those for the research fields of National Geology, Mineral Resources, Petroleum & Marine, and Geological Environment. This includes research data generated through research activities at the institute.
 - For details on the scope of research data collection related to the above, refer to the KIGAM Data Collection Guide.
- ☐ The scope of research data collection described above (research data acquired through research activities) includes DMPs.
 - A DMP is a plan for the production, preservation, management, and joint utilization of research data. It refers to a document that is prepared and submitted with the research plan when establishing a research plan for the collection, management, preservation, disclosure, and reuse of research data.

4

Research Data Storage

【 Data Storage 】

- ☐ The following activities are required in the research data storage phase:
 - Store data in compliance with relevant standards.
 - Establish a short-term retention plan to minimize potential losses.
 - Identify the number of copies of research data to be stored and synchronization methods.
 - Provide storage for research data.
 - Have a copy site for backup of data delivery systems stored in cloud-based services.
 - Provide data download service from the backup site in the event of a power outage.
 - Provide criteria for comparing storage solutions.
 - Ensure integrity and accessibility when backing up data.
- ☐ For details on backup, recovery, and preservation of research data, refer to the KIGAM Data Preservation Guide.

【 Preferred File Format 】

- ☐ Must be as machine readable as possible, considering the continuous access and potential reuse of research data.
- ☐ Non-proprietary and open formats should be used wherever possible to enhance the usability of the research data.
- ☐ The preferred format for research data files when depositing data is listed in <Table 2>³⁾.

<Table 2> Deposited data preferred format

Data type	Information about the file format	File extension
Geotechnical/ Geo-environmental data	<ul style="list-style-type: none"> • Association of Geotechnical and Geo-environmental Specialists (preferably 	<ul style="list-style-type: none"> • ags

3) NGDC. (2023.). NGDC Depositing Data - Preferred Digital Formats. Retrieved May 19, 2023 Available at: <https://www.bgs.ac.uk/download/ngdc-depositing-data-preferred-digital-formats/>

Data type	Information about the file format	File extension
	version 3.1 or 4.0)	
Geophysical data	• Log ASCII Standard	• las
	• Seismic data	• sgy
	• Sidescan sonar data	• xtf
General Scientific Data	• Microsoft Excel files	• xls
	• Comma-separated value files	• xlsx
	• Data files (with read me file on software)	• csv
	• Tab delimited data file	• dat
	• Portable document format (PDF, PDF/A)	• pdf
	• Extensible mark-up language	• xml
Text	• Microsoft Word document	• json
	• Text file (plain/ASCII)	• doc
	• Rich text format	• docx
Presentations	• Microsoft PowerPoint presentation	• ppt
Geospatial / GIS data	• ESRI shapefile/MapInfo/QGIS files/GeoTIFT	• shp
Databases	• Microsoft Access database	• aacdb
	• Oracle export	
	• MySQL export format	
	• SQLite format	
Still images	• Tagged image format	• tif
	• Joint Photographic Exports Group	• jpg
	• Portable Network Graphics	• png
	• Drawing Interchange Format (AutoCAD)	• dfx
Video	• Apple Quick Time Movie	• mov
	• Audio Video Interleaved	• avi
	• Digital Moving Picture Exchange Bitmap	• dpx
	• Moving Picture Experts Group	• mp4

【 Naming Rules 】⁴⁾

- ☐ When applying the naming method to research data, you can avoid problems such as name duplication, identity confusion, and future renaming and sorting efforts.
- ☐ KIGAM recommends that the following rules are applied.
- File and folder names should be meaningful and as descriptive as possible.
 - Files and folders should be sorted consistently.
 - Dataset names should be unique and descriptive, while reflecting the content of

4) Purdue University (n.d). File Naming Conventions. Available:
<https://guides.lib.purdue.edu/c.php?g=353013&p=2378293>

the dataset.

- Keep names short by using abbreviations where possible.
- A list of abbreviations and their descriptions, as well as other relevant information such as file type, file version information, software information, or recurring metadata associated with all files, such as data collection location, should be written to the README file.
- Use the ISO standard YYYYMMDD or YYMMDD/YYMM when entering dates.
- Choose a standard vocabulary for file names.
- When using punctuation, use the same punctuation symbols, capitalization, hyphens, and spaces.
- When using numbers, specify the number of digits to use so that the file is listed as a number.

【 Data Version Management 】

- ☐ New versions are created when changes occur in the structure, content, or status of information resources, and in the case of research data, new versions of dataset may be created when processing and modifying existing research data or adding data.
- ☐ Researchers should accurately cite the version of the dataset that supports their research results for research reproducibility and reliability.
- ☐ Data versioning refers to ensuring that a specific dataset version can be uniquely referenced to ensure the integrity and authenticity of the data.
- ☐ Revisions to dataset and metadata result in version changes, and the choice between major and minor versions is up to the research data submitter.
 - Minor version: changes to the basic information and metadata of the study data (e.g., 1.1, 1.2, ...).
 - Major version: Changes to metadata and file data (e.g., 2.0, 3.0, ...).
- ☐ Change history log records are saved for each data version change, and research data users are supported to provide change history and display changes through dataset details.
 - Within the change history log, information on change time and changer (for identification) is included.

【 Version Numbering Scheme 】

- ☐ A consistent version numbering scheme allows you to track the existence of new versions and changes to their data, and clearly distinguish between versions you've used in the past and the version you're currently working on.
- ☐ Version numbering schemes can be used by applying the method in <Table 3>⁵⁾

<Table 3> Version numbering scheme examples

Category	Content
Numbering system 1	<ul style="list-style-type: none"> • Apply a two-part numbering convention of Major.Minor (e.g., V2.1) for data archiving. • The Major part represents changes in the content or form of the dataset that may result in a change in scope, context, or purpose of use, and the Major part is renumbered when the revision occurs as follows <ul style="list-style-type: none"> - A significant number of new data items are added to or deleted from the collection. - The temporal or spatial baseline is changed, resulting in a change in data values. - Additional data attributes are introduced. - Changes to the data generation model - Data item format has changed • Minor part represents improvements in quality over existing data items and are renumbered when they are revised without affecting the purpose or scope of the initial collection (minor portions are numbered from 0). <ul style="list-style-type: none"> - Renaming data attributes - Correcting errors in existing data - Rerunning the data generation model by adjusting some parameters.
Numbering system 2	<ul style="list-style-type: none"> • Apply a two-part numbering convention of Major.Minor (e.g., V2.1) based on revision level. • Revision level 1 and revision level 2 renumber the Minor part. • Revision level 3 through revision level 4 renumber the Major

5) ARDC. (2023.). Data versioning. Retrieved May 14, 2023 Available: <https://ardc.edu.au/resource/data-versioning/>

Category	Content
	<p>part, with access to previous version data.</p> <ul style="list-style-type: none"> • Revision level 5 requires deletion of old data <ul style="list-style-type: none"> - Revision Level 1: No changes to the data itself. Changes to metadata, data file types, data access websites, correcting and updating invalid files and typos (e.g., adding new keywords) - Revision level 2: Changes to the dataset that do not significantly impact the use of the data (e.g., adding 5 missing data values) - Revision level 3: Inserting additional data collected in new time periods and locations (e.g., adding a year of data after data release) - Revision level 4: Data structure modifications, value and attribute name changes - Revision level 5: Data changes that require deletion of previous versions of the data (e.g., errors found in numbers, algorithms used for calculations)
Numbering system 3	<ul style="list-style-type: none"> • Apply a simple single number to represent data revisions and versions (e.g., V1, V2). <ul style="list-style-type: none"> - Example: Major.Minor(V2.1) (the numbering in the Minor part starts at 0)

【 Meaning of Research Data Quality Assurance 】

- ☐ This refers to ensuring that digital objects deposited conform to a variety of standards. These standards include acceptable formats, metadata schemas, metadata content, and links to other digital objects.
- ☐ It pertains to the “technical quality” of the digital object’s creation or its acquisition before deposit, rather than its “scientific quality.”

【 Scope of Application 】

- ☐ Should be implemented at various stages such as collection, entry, and verification of research data. This necessitates the development of quality control procedures at each stage.

【 Quality Control Measures 】

- ☐ Need to be executed at various stages, such as collection, input, and verification of research data. It’s essential to develop quality control procedures at each stage.
- ☐ During data review, the following quality controls should be checked:
 - Is the reviewer’s name correctly written?
 - Is the type of data review(comprehensive or spot check) accurately recorded?
 - Is the data structured and packaged in a way that enhances user accessibility?
 - Are the files readable and executable without errors?
 - Is the data machine-readable and does it utilize an open format?
 - Does the file name exclude special characters?
 - Do the file names adhere to naming conventions?
 - Does the data avoid relying on dataset names to convey information like date, geographic location, classification context and etc.?
 - If any standards are widely used in the data production area, have they been applied and documented?
 - Are the units of data values documented and appropriate?
 - Are data values valid and consistent with descriptions in metadata and other documentation?
 - Are data values free from typos?

- Are data values free from duplicate data?
- Are abbreviations and codes used consistently?
- Are there no leading or trailing spaces or tabs?
- Are there no character encoding errors?
- Are capitalization and punctuation used consistently, following standard conventions?
- Are data blank values represented using the appropriate code according to the format and content definition of the data field?
- Do documentation and processing steps include information explaining values and calculation methods?
- Does the data exclude personal and sensitive information?

【 Data Authenticity 】

- ☐ Digital information can be easily copied and altered. This requires alternatives to prove data authenticity and prevent unauthorized access.

【 Quality Control 】

- ☐ KIGAM issues DOI or IGSN through the registration API at the time of final approval of the research dataset. This is done to track and manage identification and utilization.
 - IGSN is issued for geosample dataset.
 - DOI is issued for other types of dataset other than geosample dataset.
- ☐ KIGAM became the first IGSN registrar in Asia in 2015⁶⁾. It and can assign an internationally recognized unique identification number to geosamples such as rocks and drill cores. This ensures a higher reliability of research data in the analysis of research data.

【 What to Include When Documenting Quality Controls 】⁷⁾

- ☐ Documentation of quality controls should include the following items:
 - Results of the research data quality assessment
 - The individual who performed the quality assessment, their skill requirements, and training records.
 - Methods used in the quality assessment.

6) <https://igsn.github.io/membership/>

7) USGS(n.d.). documenting-data-quality-considerations.
https://www.usgs.gov/data-management/documenting-data-quality-considerations?qt-science_support_page_related_con=0#qt-science_support_page_related_con

- Research data quality indicators used.
- Research data validation procedures.
- Methods for screening research data.

【 FAIR Principles 】

☐ Research data should adhere to the FAIR principles (Findable, Accessible, Interoperable, Reusable)

☐ The FAIR principles provide a useful framework for sharing data in a manner that maximizes its use and reuse.

☐ Findability

- Data should be assigned sufficient metadata and a unique, persistent identifier that can be easily discovered by others
- This includes assigning persistent identifiers (e.g., DOI), having sufficient metadata to describe the data, and ensuring that the data can be found through international search portals.

☐ Accessibility

- Data should be discoverable to both humans and machines through standardized communication protocols and can be authenticated and authorized where necessary.
- Data may be sensitive due to privacy concerns, national security, or commercial interests.
- If data cannot be opened, there must be clarity and transparency about the conditions governing access and reuse.

☐ Interoperability

- Relevant data and metadata use a formal, accessible, shared, and widely applicable language for knowledge representation.
- Data and metadata use community-accepted languages, formats, and vocabularies.
- Metadata should reference and describe relationships with other data, metadata, and information through identifiers.

☐ Reusability

- Relevant metadata should provide rich and accurate information, and data

should be accompanied by clear usage licenses and detailed provenance information.

- Metadata for reusable data should remain as rich and accurate as it was initially entered.
- Reusability should not be reduced for the purpose of describing a discovery in one specific publication.
- Clear machine-readable licenses and provenance information about how the data were generated are required.
- Discipline-specific data and metadata standards should be used to provide rich contextual information that allows for reuse.
- Implies principles for making research data and metadata discoverable, accessible, interoperable, and reusable.

【 FAIR Data Checklist 】⁸⁾

☐ Findable

- Making data discoverable (such as DOI or Handle) includes having rich metadata describing the data and making it discoverable through discipline-specific or national and international search portals.
- <Table 4> shows the Findable checklist.

<Table 4> Version numbering scheme examples

Checklist	Content	Score
Does the dataset have any identifiers assigned?	• No Identifier	0
	• Local Identifier	1
	• Web Address (URL)	2
	• Globally unique, citable, and persistent (e.g., DOI, PURL, ARK or Handle)	3
Is the dataset identifier included in all metadata records/files describing the data?	• No	0
	• Yes	1
How is the data described with metadata?	• The data is not described	0
	• Brief title and description	1
	• Comprehensively, but in a text-based, non-standard format	2

8) ARDC (2022). FAIR Data Self Assessment Tool. Available: <https://ardc.edu.au/resource/fair-data-self-assessment-tool/>

Checklist	Content	Score
	<ul style="list-style-type: none"> Comprehensively using a formal machine-readable metadata schema 	3
What type of repository or registry is the metadata record in?	<ul style="list-style-type: none"> The data is not described in any repository 	0
	<ul style="list-style-type: none"> Local institutional repository 	1
	<ul style="list-style-type: none"> Domain-specific repository 	1
	<ul style="list-style-type: none"> Generalist public repository 	1
	<ul style="list-style-type: none"> Data is in one place but discoverable through several registries 	2

☐ Accessible

- Making data accessible may include opening data using standardized protocols.
- If there are valid reasons, such as privacy, national security, or commercial interests, that prevent from making data open access and necessitate restriction on reuse, there should be sufficient clarity and transparency about access restrictions.
- <Table 5> shows the Accessible checklist.

<Table 5> Accessible Checklist

Checklist	Content	Score
How accessible is the data?	• No access to data or metadata	0
	• Access to metadata only	1
	• Unspecified conditional access e.g., contact the data custodian for access	2
	• Embargoed access after a specified date	3
	• A de-identified / modified subset of the data is publicly accessible	4
	• Fully accessible to persons who meet explicitly stated conditions, e.g., ethics approval for sensitive data	5
	• Publicly accessible	5
Is the data available online without requiring specialized protocols or tools once access has been approved?	• No access to data	0
	• By individual arrangement	1
	• File download from online location	2
	• Non-standard web service (e.g., OpenAPI/Swagger/informal API)	3
	• Standard web service API (e.g., OGC)	4
Will the metadata record be available even if the data is no longer available?	• Unsure	0
	• No	0
	• Yes	1

□ Interoperable

- Interoperability requires that data use community-agreed formats, languages, and vocabularies.
- Metadata must use community-agreed standards and vocabulary and include links to related information using identifiers.
- <Table 6> shows the Interoperable Checklist

<Table 6> Interoperable Checklist

Checklist	Content	Score
What (file) format(s) is the data available in?	• Mostly in a proprietary format	0
	• In a structured, open standard, non-machine-readable format	1
	• In a structured, open standard, machine-readable format	2
What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?	• Data elements not described	0
	• No standards have been applied in the description of data elements	1
	• Standardized vocabularies/ontologies/schema without global identifiers	2
	• Standardized open and universal using resolvable global identifiers linking to explanations	3
How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?	• There are no links to other metadata	0
	• The metadata record includes URI links to related metadata, data and definitions	1
	• Metadata is represented in a machine-readable format, e.g., in a linked data format such as Resource Description Framework (RDF)	2

☐ Reusable

- Reusable data should maintain the richness and accuracy of the initial metadata entered.
- Domain-specific data and metadata standards should be used to provide rich, contextual information that can be accurately interpreted and reused.
- <Table 7> shows the Reusable Checklist

<Table 7> Reusable Checklist

Checklist	Content	Score
Which of the following best describes the license/usage rights attached to the data?	• No license	0
	• Non-standard text-based license	1
	• Non-standard machine-readable license (clearly indicating under what conditions the data may be reused)	2
	• Standard text-based license	2
	• Standard machine-readable license (e.g., Creative Commons)	3
How much provenance information has been captured to facilitate data reuse?	• No provenance information is recorded	0
	• Partially recorded	1
	• Fully recorded in a text format	2
	• Fully recorded in a machine readable format	3

【 Research Data Preservation 】

- ☐ KIGAM has developed and complies with the guide for the preservation of research data, and the details of the preservation of research data are described in the guidelines.
 - KIGAM Data Preservation Guide includes the concept of research data preservation, selection and evaluation of preserved data and repositories, etc.
- ☐ Data Management Guide focuses on metadata description for the preservation of research data.

【 Research Data Preservation and Description 】

- ☐ The research data preservation phase requires the following activities:
 - Measure to preserve and retain the administrative characteristics of the data for the long term.
 - Develop a plan to preserve the data for the long term.
 - Determine what data to preserve, where to preserve it, and what documentation is needed with the data.
 - Create metadata and documentation for preservation.
 - Organize and store the data.

【 Metadata Meaning 】

- ☐ Metadata refers to data used to describe data, consisting of its title, the producer of the data, the equipment and method of producing the data, the content of the data, the location (location coordinates) and time of acquisition, the format of the data, and the quality of the data. All these elements are necessary for describing the data adequately.

【 Considerations for Metadata Description 】

- ☐ Adhere to the metadata schema defined by KIGAM.
- ☐ Provide rich descriptions for long-term preservation and utilization of research data.

- ☐ Identify how metadata records were created.
- ☐ Identify which metadata standards were used.
- ☐ Identify what tools were used.
- ☐ Ensure that records are created at the beginning of the project and updated as research progresses.
- ☐ Consider where to deposit the metadata.
- ☐ Consider community standards when deciding on metadata standards and repositories.

【 Metadata Descriptions 】 ⁹⁾

- ☐ Metadata used during collection and deposition uses the Dublin Core (DC) format.
 - The 15 DC metadata elements follow the KIGAM Data Collection Guide

⁹⁾ National Library of Korea. DCMI DCMES Features.
<https://librarian.nl.go.kr/LI/contents/L10202000000.do>

【 Meaning of 】

- ☐ Research data must be discoverable and accessible for reuse.
- ☐ Research data repositories should provide features that make it easy for users to search for and access research data.
- ☐ Repositories allow digital objects to be reused over time, ensuring that reliable information is available to support understanding and use.
- ☐ Searchable data may not have unrestricted access, so there may exist confidentiality, reuse permissions, access restrictions, licenses, etc.

【 Research Data Disclosure and Sharing 】

- ☐ Researchers (data producers) can select private, internal, or external disclosure when submitting research data, and can specify the time limit for disclosure through Embargo.
- ☐ Researchers are required to make research data available to as many people as possible while protecting research confidentiality and as permitted by applicable regulations.
- ☐ When research data is published or shared, it should be organized according to a common standard, and accompanied by a license with appropriate terms so that it can be easily reused by other researchers.
- ☐ Researchers should protect and manage sensitive data containings personal information by transforming it into a de-identified form before releasing or sharing it.

【 Importance and Necessity of Sharing Research Data 】¹⁰⁾

- ☐ Recycling and management of research data are becoming increasingly important

10) DataON. Research data management and utilization guidelines. Available:
https://dataon.kisti.re.kr/data_mgnt_guideline_08.do

as we move towards data-intensive research, where scientific discoveries are made by using vast amounts of data produced by various measurement and experimental instruments.

- ☐ With the open science and open access movements, data and publications produced by publicly funded research projects are increasingly being published in public repositories.
- ☐ Sharing research data increases the availability of data to other researchers or institutions.
- ☐ Sharing research data increases the reputation of individual researchers or research institutions through data citation, etc., and provides opportunities to develop better research through data verification, etc.
- ☐ Sharing data can reduce the cost of duplicate production and duplicate publication of data, which can be used to focus on future research.
- ☐ Most pre-processed data can be shared if sensitive issues such as privacy and research ethics are well considered.
- ☐ When R&D institutions promote research data disclosure through DMP, standardization of data collection, storage, and metadata creation is the best way to manage research data to realize research data sharing.
- ☐ Details on data repositories for sharing and managing research data are specified in the KIGAM Data Preservation Guide.
 - KIGAM Data Preservation Guide includes definitions of data repositories, considerations for selection, examples of repositories, etc.

【 FAIR Principles: Reusable 】

- ☐ The FAIR Principles provide a useful framework for thinking about reusing and sharing research data as much as possible, and the Reusable section provides a detailed checklist for sharing and utilizing research data.
- ☐ Below are more details on the Reusable section:
 - Metadata about research data should be rich and accurate, and research data should be accompanied by a clear usage license and detailed provenance

information.

- Reusable research data should not be summarized and should retain the richness of the raw data.
- Clear machine-readable licenses and provenance information about how the data was created are required.
- Discipline-specific data and metadata standards should be used to provide rich contextual information that allows for reuse.
- Multiple accurate and relevant attributes should be provided to enable reuse.
- Both research data and metadata should be machine-readable and conform to community standards.
- Metadata should include information about the license under which the research data can be reused.
- Metadata should reference a standard reuse license, which means a reuse license that is machine understandable.
- Metadata should include provenance information according to community-specific standards and cross-community languages and should be in a machine-readable format.

【 Embargo 】¹¹⁾

- ☐ A period of time during which descriptions of the dataset, such as title, author, metadata, and abstract, are available, but public access to the dataset is restricted until a time specified by the researcher has elapsed. At the end of the embargo period, the dataset is made available through public or mediated (data available after application approval) access¹²⁾.
- ☐ The data repository is obligated to manage the embargo information set by the researcher.
- ☐ When depositing a dataset in embargo, the embargo start and end dates must be provided. The data repository restrict access to the dataset until the end of the embargo period, and users are able to use the dataset after the embargo end date.

11) Research data management and utilization for government-funded research institutes in Korea.
Available:
<https://www.nst.re.kr/www/selectBbsNttView.do;JSESSIONID=0C1DA6F31EBF881198E95FF8105B74AB?key=64&bbsNo=34&nttNo=2205&searchCtgr=&searchCnd=all&searchKwd=&integrDeptCode=&pageIndex=5>

12) Curtin University (n.d.). Embargoes.
Available:<https://researchtoolkit.library.curtin.edu.au/data/research-data-management/publication/#s-lg-box-21330046>

- ☐ Researchers may be asked to provide a rationale for applying an embargo to a dataset, as this will restrict user access to the dataset.
- ☐ Examples of situations where an embargo may be applied to a dataset include:
 - When publishing research based on a dataset, prior to making the dataset accessible.
 - When you need to finalize commercial interests from research data prior to publication, such as a patent application.
 - When the dataset includes sensitive data.

【 Data Citation 】

- ☐ Data citation means that researchers provide references to their data in the same way that they provide bibliographic references to research outputs such as scholarly articles and reports in research publications.
- ☐ Research data citation is recognized as one of the key practices for recognizing data as a major research output.

【 Necessity of Citing Research Data 】

- ☐ Citing dataset provides rewards (acknowledgment) to research data creators and research data stewards.
- ☐ Provides appropriate credit to researchers, rewards academic endeavors, and gives credit to research data managers and repositories that manage research data over the long term.
- ☐ Citation of research data holds creators and custodians accountable for their research data and reduces the risk of plagiarism when data are properly cited.
- ☐ Research data citation enables data discovery, reuse, and validation of research results, which improves the discoverability and reusability of research data.
- ☐ When using published research data, copyright and license terms specified by the research data repository must be identified and followed.
- ☐ Citation of research data creates a formal system of recognition and reward for research data producers in the scientific community.

- ☐ Citation of research data enables the tracking and calculation of research data, which in turn enables the evaluation of the impact of research data such as impact factors.
- ☐ Citation of research data not only increases the transparency of research data production in publications, but also encourages the production of high-quality research data.
- ☐ Citation of research data becomes routine, ensuring that research data is reproducible with the highest quality research results.
- ☐ It is necessary to measure the impact factor, in a similar way to academic papers.
- ☐ The use of persistent identifiers within citations is key to ensuring that all research data citations are accurately tracked and counted.
- ☐ Citation indexes measure the reuse of research data, and these metrics can be commonly used for performance evaluation and reporting.

【 Components of Data Citation 】

- ☐ Data citations have the following components
 - Author
 - Title
 - Publication date
 - Version or edition number
 - Distributor
 - Persistent identifier
 - Access date and time

【 Data Identifiers 】

- ☐ Data identifiers are essential in all computer-based systems and allow scholarly records to be cited.
- ☐ Persistent identifiers are identifiers that uniquely identify all kinds of objects and require constant management and updating.

- ☐ A persistent identifier that is often used to cite research data is the DOI.
- ☐ Benefits of DOI¹³⁾
 - Ensures a high level of confidence in quality and accuracy.
 - Supported by the International DOI Federation (IDF) and registrar infrastructure.
- ☐ Data citation
 - DOI requires metadata elements that produce clear data citations.
 - Example: Hanigan, Ivan (2012): Monthly drought data for Australia 1890–2008 using the Hutchinson Drought Index. The Australian National University Australian Data Archive. (doi. org/10.4225/13/50bbfd7e6727a)
 - Permanence
 - DOI indicates that the dataset is well maintained and available for long-term use
 - Accessibility
 - DOI facilitates online accessibility to research data

【 How to Cite Data 】

- ☐ Published research data should be cited in the same way as other academic outputs.
- ☐ When publishing and citing data registered and collected in the research data platform, the platform data citation must be specified.
- ☐ Just as there are various citation styles and formats for academic articles, there are also various citation styles for data.
- ☐ Data citations typically consist of the following elements: author, publication date, data title, version or edition number, archive or distributor, persistent identifier, and date and time of access.
- ☐ A standard citation contains the elements illustrated in <Figure 2>.

13) ARDC (2022). ARDC DataCite DOI service. Available:
<https://ardc.edu.au/services/ardc-identifier-services/ardc-datacite-doi-service/>

Author. (Year of publication). Title. Version*. Publisher. DOI*.

<Figure 2> Research Data Standard Citation Format (* is optional)

- ☐ Data citation methods for the Geo Big Data Open Platform are as follows¹⁴⁾
 - If the recipient of the research data discloses the results of its utilization in an article, report, or website, the recipient must include a data citation in the result.
 - Data citation format
 - Author. (year of data production). Dataset name. Data publishing organization. DOI
 - Example: Hong Gil-dong. (2020.11.01.). Infrared spectroscopy data. Unknown. Retrieved from <http://doi.org/----->
- ☐ The terms and conditions for citing data from the Geo Big Data Open Platform are specified in the KIGAM Data Management Rules (Chapter 6, Open and Use Conditions of Research Data), and are as follows
 - The individual using the research data must acknowledge the contribution of those who produced the research data. If not otherwise specified in the research data license, they should formally specify the use of the research data using a citation.
 - Anyone using research data should check the license terms and conditions if a license has been applied to the research data. They should comply with the terms of use including acknowledging, authorship, sources, etc.

【 Data Publication 】

- ☐ Publication Procedures
 - Data publication is the act of making research data publicly available for (re)use by other researchers. Data can be published online in an accessible form through data repositories such as research institutions.
 - Separate quality check procedures can be introduced, such as peer review by experts in related subject areas or data experts.
- ☐ Assigning data identifiers

14) KIGAM research report 「Establishment of National Geoscience Data Center(NGDC) foundation through the geo big data open platform development」 Appendix 4: Guidelines for Opening and Providing Research Data

- Identifiers such as DOI and IGSN are assigned to publicly available research data in order to identify dataquest and track their utilization.

☐ Apply licenses

- The Geo Big Data Open Platform provides Creative Commons license standards and support services for research data disclosure and copyright protection. Users can select and apply appropriate licenses for each research dataset.
- For other contents, refer to the KIGAM Ethics • Copyright • Licensing Guidelines.

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