

---

# 연구데이터 관리 지침

---

2023. 5.

**KIGAM** 한국지질자원연구원  
지질자원데이터센터

# 목 차

1. 개요 .....	1
2. 연구데이터 관리체계 .....	2
3. 연구데이터 수집 .....	8
4. 연구데이터 저장 .....	9
5. 연구데이터 품질보장 .....	14
6. 연구데이터 기술(Description) .....	21
7. 연구데이터 공유 및 활용 .....	23

## 표 목차

〈표 1〉 데이터 생애주기 구성 요소 세부 내용 .....	5
〈표 2〉 기탁 데이터 선호 형식 .....	9
〈표 3〉 버전 넘버링 체계 예시 .....	12
〈표 4〉 Findable 체크리스트 .....	17
〈표 5〉 Accessible 체크리스트 .....	18
〈표 6〉 Interoperable 체크리스트 .....	19
〈표 7〉 Reusable 체크리스트 .....	20

## 그림 목차

〈그림 1〉 연구데이터 관리를 위한 데이터 생애주기(Data Lifecycle) .....	5
〈그림 2〉 연구데이터 표준 인용 형식 .....	28

# 1

## 개요

### 【 목 적 】

- ☐ 본 문서의 목적은 한국지질자원연구원 지질자원데이터센터에서 연구데이터 관리 시 적용할 수 있는 연구데이터 관리 지침(안) 제시

### 【 대 상 】

- ☐ 지질자원데이터센터 연구데이터 관리 담당자

### 【 적용범위 】

- ☐ 원내에서 연구 활동을 통해 생산된 연구데이터와 외부 기관 및 개인으로부터 기탁된 연구데이터에 적용함

### 【 준 용 】

- ☐ 본 지침에서 정하지 아니한 사항은 국가과학기술연구회의 연구데이터 관리 가이드라인 및 한국지질자원연구원의 제·규정 준용 가능

## 【 의 미 】

- ☐ 연구데이터 관리는 연구 프로젝트를 수행하는 동안에 생성되고 수집된 데이터의 저장, 공유, 접근, 보존 및 재사용의 방법을 다루는 것을 의미함
- ☐ 이 외에도 데이터관리계획(Data Management Plan, 이하 DMP), 프로세스 그리고 장기 보관 및 공유가 포함됨
- ☐ 연구데이터의 정의는 연구데이터 수집 지침의 내용을 따름
- ☐ 한국지질자원연구원의 국가연구개발사업에서 생산, 보유하는 모든 연구데이터를 관리 대상으로 하며, 공개 대상 데이터는 보안규정 및 한국지질자원연구원의 제·규정 등을 고려하여 선별함

## 【 목적 및 목표 】

- ☐ 연구데이터 관리의 목적은 장기적인 접근 및 지속이 가능한 데이터 활용을 보장함으로써 연구데이터 공유를 촉진하고 향후에 연구데이터를 재사용할 수 있도록 하는 것임
- ☐ 연구데이터의 재사용이 원활히 이루어지기 위해서는 프로젝트에 직접 참여하지 않은 사람에게 데이터를 제공하더라도 해당 데이터를 이해하고 이를 적절히 사용할 수 있도록 데이터를 생성해야 함
  - 이를 위해 DMP를 작성하고 기본적인 데이터 설명과 관리 및 재사용 방법 등을 공유하는 것임
- ☐ 연구데이터 관리를 통한 기대효과로는 다음과 같은 항목을 들 수 있음
  - 연구데이터 손실 위험 감소
  - 연구데이터 제어의 효율성 향상 및 용이성 보장
  - 연구데이터 가시성 향상을 통한 인용 및 향후 협업 증가
  - 연구 무결성 입증 및 연구 결과 검증
  - 지식 이전을 통한 연구 영향력 증대
  - 연구데이터 재사용으로 인한 연구 발전

## 【 원 칙 】

- ☐ 데이터는 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙을 준용함

- Findable(검색 가능): 데이터에 메타데이터와 디지털 객체 식별자(Digital Object Identifier, 이하 DOI) 등 영구 식별자를 할당하고 연구데이터 리포지터리에 등록하여 검색 가능하도록 함
- Accessible(접근 가능): 표준화된 프로토콜을 사용하여 식별자를 통해 데이터 연구데이터가 이용가능하지 않을 경우 메타데이터에 접근 가능하도록 함
- Interoperable(상호 운용 가능): 데이터는 표준적인 방식과 어휘를 사용하여 상호 운용 가능하도록 해야 하며, 관련된 다른 데이터에 대한 인용을 포함함
- Reusable(재사용 가능): 명확한 데이터 라이선스와 프로비넌스 정보를 제공하여 재사용 가능하도록 함

## 【 데이터 관리 체계 수립 】

### □ 정책과 절차

- 한국지질자원연구원은 연구데이터관리규정을 통해 연구자가 책임 있는 연구 수행과 데이터 관리를 위해 연구데이터 관리 기준을 정의하고 있으며, 데이터 관리에 필요한 정책과 절차를 수립하여 기관 내에 공표하고 있음

### □ 전담조직 및 인력

- 한국지질자원연구원에서는 연구데이터 관리 및 활용 지원을 위한 전문 인력을 충원하고 전담조직인 연구데이터 담당부서를 구성하여 운영하고 있으며, 한국지질자원연구원 연구데이터관리 규정(제 3장 연구데이터 관리부서)에 명시되어 있음
  - 한국지질자원연구원은 연구데이터관리를 총괄 · 조정하고 연구데이터의 수집 · 관리 · 보존 · 개방 및 활용에 관한 제반업무 수행을 위하여 연구데이터 담당부서를 운영하여야 하며, 연구데이터 담당부서를 주관부서로 하고 있음
  - 주요 업무는 아래와 같음
    - 연구데이터관리에 관한 기본 정책의 수립 및 제도의 개선
    - 연구데이터관리체계 구축 및 운영
    - 연구데이터의 표준화, 품질관리, 보존, 개방 및 활용기술의 개발 · 연구 및 보급
    - 연구데이터관리위원회 운영
    - 연구데이터관리에 관한 교육훈련과 상담 및 기술지원
    - 연구데이터관리를 위한 다른 기관과의 연계 · 교류 · 협력에 관한 사항
    - 그 밖에 연구데이터관리에 필요한 사항
- 한국지질자원연구원은 연구데이터관리위원회를 운영하여 DMP 및 연구데이터 관련 업무 수행에 관한 사항을 심의하고 있으며, 한국지질자원연구원 연구데이터관리규정(제2장 연구데이터관위원회)에 명시되어 있음
  - 각 부서장을 포함하여 25명 이내로 구성하며, 위원장은 부원장으로 함. DMP 심

의와 데이터관리의 효율성을 위하여 특별한 사유가 없는 한 위원은 연구업무심의회 위원과 동일하게 구성함

□ 요구사항 정의

- 연구데이터 관리를 위해서는 한국지질자원연구원에서 생산 또는 보유하고 있는 연구데이터의 크기, 종류, 생산주기, 검색 요구사항을 정의하고 있음
- 체계적인 관리와 폭넓은 공유·활용을 위해 연구데이터 및 프로세스의 표준화 방안을 마련하고 있음

□ 연구데이터 인프라

- 한국지질자원연구원에서는 안정적인 연구데이터 관리와 장기간 보존을 제공하고 다양한 연구데이터 공유 활용을 지원하는데 필수적인 IT 인프라를 제공하고 있음

□ 지원 서비스

- 한국지질자원연구원에서는 체계적인 연구데이터 관리·활용을 위해 연구자에게 교육, 자문, 지원 서비스 등을 제공하고 있음



【 데이터 생애주기(Data Lifecycle)에 따른 관리 방법 】

- 본 지침의 연구데이터 관리 방법은 데이터 생애주기(Data Lifecycle)를 따르고 있으며, 데이터 생애주기(Data Lifecycle)는 다음 <그림 1>과 같음<sup>1)</sup>



<그림 1> 연구데이터 관리를 위한  
데이터 생애주기(Data Lifecycle)

- 다음 <표 1>은 데이터 생애주기 구성 요소의 세부 내용을 기술한 것임

<표 1> 데이터 생애주기 구성 요소 세부 내용

단계	구성 요소	세부 내용
수집 (Collect)	DMP 작성	<ul style="list-style-type: none"> <li>• 생성하거나 수집해야 하는 데이터 결정</li> <li>• 데이터, 메타데이터의 형식 및 품질에 대한 표준 식별</li> <li>• 데이터 관리자의 역할 및 책임</li> <li>• 데이터 공유 계약 개발</li> <li>• 데이터 접근 제어 및 저작권 이해</li> <li>• 기존 데이터 탐색 및 확보</li> <li>• 데이터 수집(실험, 관측, 측정, 시뮬레이션)</li> </ul>
저장 (Store)		<ul style="list-style-type: none"> <li>• 해당 데이터의 메타데이터 및 문서화 기술</li> <li>• 관련 표준을 준수하여 데이터 저장</li> <li>• 잠재적 손실 최소화를 위한 단기 보존 계획 수립</li> <li>• 저장할 연구데이터 사본의 개수와 동기화 방법 제시</li> </ul>

1) 김주섭, 김선태, 전예린 (2019). 연구 데이터 관리를 위한 데이터 생명주기(Data Lifecycle) 제안. 한국문헌정보학회지, 53(4), 309-340.

단계	구성 요소	세부 내용
관리 (Manage)		<ul style="list-style-type: none"> <li>연구데이터 저장을 위한 스토리지 제공</li> <li>클라우드 기반 서비스에 저장된 데이터 전송 시스템의 백업을 위한 사본 사이트 보유</li> <li>정전 시 백업 사이트에서 데이터 다운로드 서비스 제공</li> <li>스토리지(저장 공간) 솔루션 비교 기준 제시</li> <li>데이터 백업 시 무결성 및 접근성 보장</li> </ul>
		<ul style="list-style-type: none"> <li>데이터의 무결성을 유지하면서 진본, 신뢰성 및 가용성 보장</li> <li>허용 가능한 데이터 구조 또는 파일 형식 보장</li> <li>품질 보증 및 관리</li> <li>데이터 품질 문서화</li> <li>데이터 백업</li> </ul>
	품질보장	
	보안	<ul style="list-style-type: none"> <li>공개 위험 제한 원칙</li> <li>이용이 제한된 데이터 수집 방법</li> <li>접근 제한(개인 정보 보호 및 기밀성 보호)</li> <li>데이터 개인정보 익명화</li> <li>우발적인 데이터 손실, 손상 및 무단 접근으로부터 데이터 보호</li> </ul>
접근 및 이용 (Access & Use)		<ul style="list-style-type: none"> <li>지정된 이용자와 재사용자 모두가 매일 데이터에 액세스 할 수 있는지 확인</li> <li>접근 제어 및 인증 절차 적용</li> <li>선호 파일 형식 제시</li> <li>프로젝트 사이트에 프로젝트 등록 및 오픈 리포지터리에 데이터 저장</li> <li>검색, 분석, 용도 변경, 배포, 표현 방법 지원</li> </ul>
보존 (Preserve)		<ul style="list-style-type: none"> <li>데이터의 관리 특성을 장기간 보존하고 보유하기 위한 조치</li> <li>데이터를 장기적으로 보존할 계획 수립</li> <li>보존할 데이터, 보존할 위치 및 데이터와 함께 필요한 도큐멘테이션 결정</li> <li>보존을 위한 메타데이터 및 도큐멘테이션 생성</li> <li>데이터 구성 및 저장</li> </ul>
공유 및 활용 (Share & Reuse)	인용	<ul style="list-style-type: none"> <li>데이터셋 및 프로젝트 도큐멘테이션 기술</li> <li>데이터 이력 문서화</li> <li>데이터의 이유, 사용자, 시기, 위치 및 방법을 설명하여 데이터 문서화</li> </ul>

단계	구성 요소	세부 내용
	출판	<ul style="list-style-type: none"> <li>• 카탈로그 작성 및 설명</li> <li>• 배포된 데이터에 대한 일관성 있는 식별 보장</li> <li>• 데이터 공개 프로세스</li> <li>• DOI, IGSN<sup>2)</sup>((International Generic Sample Number, 이하 IGSN) 발급</li> <li>• 데이터 배포 및 공유</li> <li>• 저작권 설정</li> <li>• 데이터에 대한 정보 배포, 데이터 홍보</li> </ul>

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

## 【 연구데이터 수집 】

- 한국지질자원연구원은 연구데이터 수집 지침을 작성하여 이를 준수하고 있으며, 연구데이터 수집에 대한 자세한 사항은 해당 지침에 기술되어 있음
  - 연구데이터 수집 지침에는 연구데이터의 정의, 수집범위, 선택 및 평가, 기탁, DMP 작성 구성 및 예시 등의 내용이 포함됨

## 【 DMP 작성 안내 】

- 연구데이터 수집 단계에서는 DMP 작성과 더불어 다음과 같은 활동이 요구됨
  - 생성하거나 수집해야 하는 데이터 결정
  - 데이터, 메타데이터의 형식 및 품질에 대한 표준식별
  - 데이터 관리자의 역할 및 책임
  - 데이터 공유 계약 개발
  - 데이터 접근 제어 및 저작권 이해
  - 기존 데이터 탐색 및 확보
  - 데이터 수집(실험, 관측, 측정, 시뮬레이션)
  - 해당 데이터의 메타데이터 및 도큐멘테이션 기술
- 한국지질자원연구원은 국토지질, 광물자원, 석유해저, 지질환경 등 지질자원 데이터셋을 수집 및 관리하며 이는 연구원에서 연구 활동을 통해 획득한 연구데이터를 포함함
  - 위 내용과 관련된 연구데이터 수집범위에 대한 자세한 사항은 연구데이터 수집 지침을 따름
- 상기 명시된 연구데이터 수집 범위(연구 활동을 통해 획득한 연구데이터)에는 DMP가 포함됨
  - DMP는 연구데이터의 생산·보존·관리 및 공동 활용 등에 관한 계획으로, 연구데이터의 수집·관리·보존·개방 및 활용을 위하여 연구계획수립 시 연구계획서와 함께 작성하여 제출하는 문서를 의미함

## 【 연구데이터의 저장 】

- ☐ 연구데이터 저장 단계에서는 다음과 같은 활동이 요구됨
- 관련 표준을 준수하여 데이터 저장
  - 잠재적 손실 최소화를 위한 단기 보존 계획 수립
  - 저장할 연구데이터 사본의 개수와 동기화 방법 제시
  - 연구데이터 저장을 위한 스토리지 제공
  - 클라우드 기반 서비스에 저장된 데이터 전송 시스템의 백업을 위한 사본 사이트 보유
  - 정전 시 백업 사이트에서 데이터 다운로드 서비스 제공
  - 스토리지(저장공간) 솔루션 비교 기준 제시
  - 데이터 백업 시 무결성 및 접근성 보장
- ☐ 연구데이터의 백업 및 복구, 보존에 대한 자세한 사항은 한국지질자원연구원 연구데이터 보존 지침을 따름

## 【 선호 파일 형식 】

- ☐ 연구데이터의 지속적인 접근과 잠재적인 재사용을 고려하여 가능한 기계 가독형(machine readable)이어야 함
- ☐ 연구데이터에 대한 활용성을 제고하기 위해 가능한 독점적이지 않고 공개된 포맷(open format)을 사용해야 함
- ☐ 데이터 기탁 시의 연구데이터 파일은 <표 2><sup>3)</sup>와 같은 파일 형식(format)을 선호함

&lt;표 2&gt; 기탁 데이터 선호 형식

데이터 유형	권장 형식	파일확장자
지질공학/환경 데이터	• Association of Geotechnical and Geo-environmental Specialists(preferably version 3.1 or 4.0)	• ags
지구물리 데이터	• Log ASCII Standard	• las
	• Seismic data	• sgy
	• Sidescan sonar data	• xtf
일반 과학 데이터	• Microsoft Excel files	• xls •xlsx

데이터 유형	권장 형식	파일 확장자
	• Comma-separated value files	• csv
	• Data files (with read me file on software)	• dat
	• Tab delimited data file	
	• Portable document format(PDF, PDF/A)	• pdf
	• Extensible mark-up language	• xml • json
텍스트	• Microsoft Word document	• doc • docx
	• Text file(plain/ASCII)	• txt
	• Rich text format	• rtf
프레젠테이션	• Microsoft Powerpoint presentation	• ppt
GIS/공간정보	• ESRI shapefile/MapInfo/QGIS files/GeoTIFT	• shp
데이터베이스	• Microsoft Access database	• aacdb
	• Oracle export	
	• MySQL export format	
	• SQLite format	
이미지	• Tagged image format	• tif
	• Joint Photographic Exports Group	• jpg
	• Portable Network Graphics	• png
	• Drawing Interchange Format(AutoCAD)	• dfx
비디오	• Apple Quick Time Movie	• mov
	• Audio Video Interleaved	• avi
	• Digital Moving Picture Exchange Bitmap	• dpx
	• Moving Picture Experts Group	• mp4

#### 【 네이밍 규칙 】<sup>4)</sup>

- ☐ 연구데이터의 이름 지정 방식(naming) 적용 시 이름 중복, 정체성 혼란, 향후 이름 변경 및 정렬 작업 등에서 발생할 수 있는 문제를 예방할 수 있음
- ☐ 다음과 같은 규칙을 적용할 것을 권장함
  - 파일과 폴더의 이름은 의미가 있어야 하며, 가능한 한 명료해야 함
  - 파일과 폴더는 일관성을 유지하며 정렬되어야 함
  - 데이터셋의 이름은 해당 데이터셋의 내용을 반영하면서도, 고유하며 설명적(descriptive)이어야 함
  - 가능한 약어를 사용하여 이름을 짧게 유지함
  - 약어 목록과 그에 대한 설명파일 유형(file type), 파일 버전 정보, 소프트웨어 정보 또는 데이터 수집 위치와 같은 모든 파일과 관련된 반복되는 메타데이터 등

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

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

기타 관련 정보를 read me 파일에 작성하여야 함

- 날짜 입력 시 ISO 표준 YYYYMMDD 또는 YYMMDD/YYMM 사용하여야 함
- 파일 이름에 관한 표준 어휘 선택하여야 함
- 구두점을 사용할 시 같은 구두점 기호, 대문자, 하이픈 및 공백을 사용하여야 함
- 숫자를 사용할 시 파일이 숫자로 나열되도록 사용할 자릿수 지정하여야 함

## 【 데이터 버전 관리 】

- ☐ 정보자원의 구조, 내용 또는 상태의 변화가 발생할 때 새로운 버전이 생성되며, 연구데이터의 경우 기존 연구데이터의 가공 및 수정 또는 데이터 추가 시에 새로운 버전의 데이터셋이 생성될 수 있음
- ☐ 연구자는 연구 재현성과 신뢰도를 위해 연구 결과를 뒷받침하는 데이터셋 버전을 정확하게 인용하여야 함
- ☐ 데이터 버전 관리는 데이터의 무결성과 신뢰성을 보장하기 위하여 특정 데이터셋 버전을 고유하게 참조할 수 있도록 하는 것을 말함
- ☐ 데이터셋 및 메타데이터의 수정 시 버전 변경이 발생하며, 메이저(major) 및 마이너(minor) 버전의 여부는 연구데이터 제출자가 선택함
  - 마이너 버전: 연구데이터 기본 정보 및 메타데이터 변경(예: 1.1, 1.2 ...)
  - 메이저 버전: 메타데이터 및 파일 데이터 변경(예: 2.0, 3.0 ...)
- ☐ 각 데이터의 버전 변경 시 변경이력 로그 기록이 저장되며, 연구데이터 이용자에게는 데이터셋의 상세정보를 통해 변경 이력 제공 및 변경내용 표시가 지원됨
  - 변경이력 로그 내에는 변경시간 및 변경자(신원확인용)에 대한 정보가 포함됨

## 【 버전 넘버링 체계 】

□ 일관된 버전 넘버링 체계(version numbering scheme)를 통해 새 버전의 존재와 해당 데이터의 변경 여부를 추적할 수 있으며, 이전에 사용한 버전과 현재 작업 중인 버전을 명확하게 구분할 수 있음

□ 버전 넘버링 체계는 다음 <표 3><sup>5)</sup>의 방법을 활용하여 사용할 수 있음

<표 3> 버전 넘버링 체계 예시

항목	내용
넘버링 시스템 1	<ul style="list-style-type: none"> <li>• 데이터 버저닝을 위해 Major.Minor(예: V2.1)의 두 부분으로 된 넘버링 규칙을 적용함</li> <li>• Major 부분은 범위, 상황 또는 사용 목적의 변화를 일으킬 수 있는 데이터셋의 내용 또는 형태의 변경을 나타내며 다음과 같이 개정될 경우 Major 부분의 번호를 갱신함 <ul style="list-style-type: none"> <li>- 컬렉션에 상당한 양의 새로운 데이터 항목이 추가되거나 삭제됨</li> <li>- 시간 또는 공간 기준선(spatial baseline)이 변경되어 데이터 값이 변경됨</li> <li>- 추가적인 데이터 속성(data attributes)이 도입됨</li> <li>- 데이터 생성 모델(data generation model)의 변경</li> <li>- 데이터 항목 포맷이 변경됨</li> </ul> </li> <li>• Minor 부분은 기존 데이터 항목보다 품질이 향상된 것을 나타내며, 초기 컬렉션의 사용 목적 또는 범위에 영향을 미치지 않고 다음과 같이 개정될 경우 Minor 부분의 번호를 갱신함 (Minor 부분의 번호는 0부터 시작함) <ul style="list-style-type: none"> <li>- 데이터 속성의 이름 변경</li> <li>- 기존 데이터의 에러 수정</li> <li>- 일부 매개 변수 조정하여 데이터 생성 모델 재실행</li> </ul> </li> </ul>
넘버링 시스템 2	<ul style="list-style-type: none"> <li>• 개정 수준 레벨에 따른 Major.Minor(예: V2.1)의 두 부분으로 된 넘버링 규칙을 적용함</li> <li>• 개정 레벨 1과 개정 레벨 2는 Minor 부분의 번호를 갱신함</li> <li>• 개정 레벨 3부터 개정 레벨 4까지 Major 부분의 번호를 갱신하며, 이전 버전 데이터에 접근 가능함</li> <li>• 개정 레벨 5는 이전 데이터의 삭제가 요구됨 <ul style="list-style-type: none"> <li>- 개정 레벨 1: 데이터 자체 변경 없음. 메타데이터, 데이터 파일 유형, 데이터 접근 웹 사이트 변경, 잘못된 파일 및 오타</li> </ul> </li> </ul>

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



항목	내용
	<p>수정 및 업데이트(예: 새 키워드 추가)</p> <ul style="list-style-type: none"> <li>- 개정 레벨 2: 데이터 사용에 큰 영향이 없는 데이터셋 변경 (예: 누락된 5개 데이터 값 추가)</li> <li>- 개정 레벨 3: 새로운 기간, 장소에서 수집한 추가 데이터 삽입(예: 데이터 공개 이후 1년 간의 데이터 추가)</li> <li>- 개정 레벨 4: 데이터 구조 수정, 값 및 속성명 변경</li> <li>- 개정 레벨 5: 이전 버전 데이터의 삭제가 필요할 정도의 데이터 변경(예: 숫자, 계산에 사용되는 알고리즘에 오류가 발견된 경우)</li> </ul>
넘버링 시스템 3	<ul style="list-style-type: none"> <li>• 단순한 하나의 숫자를 적용하여 데이터의 개정 및 버전을 나타냄(예: V1, V2)</li> <li>- 예시: Major.Minor(V2.1)(Minor 부분의 번호는 0부터 시작함)</li> </ul>

## 【의미】

- ☐ 기탁된 디지털 객체가 허용 가능한 형식, 메타데이터 스키마, 메타데이터 콘텐츠 및 다른 디지털 객체에 대한 링크를 포함하여 다양한 표준 기준을 준수하도록 하는 것을 의미함
- ☐ 디지털 객체 생성 또는 보관 전 수집에서의 ‘과학적 품질’ 보다는 ‘기술적 품질’ 과 관련이 있음

## 【범위】

- ☐ 연구데이터의 수집, 입력, 확인 등 다양한 단계에서 수행되어야 하며, 해당 단계에서의 품질 관리 절차 개발이 필요함

## 【품질 관리 조치】

- ☐ 연구데이터의 수집, 입력, 확인 등 다양한 단계에서 수행되어야 하며, 해당 단계에서의 품질 관리 절차 개발이 필요함
- ☐ 데이터 검토 시에는 다음과 같은 품질 관리 사항이 확인되어야 함
  - 검토자명(name)이 정확히 기입되었는가?
  - 데이터 검토 유형이 종합 점검 또는 현장 점검을 통하여 정확하게 기입되었는가?
  - 데이터는 이용자의 접근성을 향상하는 방식으로 구조화되고 패키지화되었는가?
  - 파일이 오류 없이 읽기 및 실행할 수 있는가?
  - 데이터를 기계가 읽을 수 있는가? 개방형 형식을 사용하는가?
  - 파일 이름에 특수 문자를 포함하지 않는가?
  - 파일 이름이 네이밍 규칙을 사용하는가?
  - 데이터는 날짜, 지리적 위치, 분류 맥락 등을 전달할 때 데이터셋 이름에 의존하지 않는가?
  - 어떠한 표준이 데이터 생산 영역에서 널리 사용되는 경우, 표준을 사용하고 문서화하였는가?
  - 데이터 값의 단위는 문서화되고 적절한가?
  - 데이터 값이 유효하고 메타데이터 및 기타 문서의 설명과 일치하는가?
  - 데이터 값에 오타가 없는가?
  - 중복 데이터가 없는가?
  - 약어 및 코드가 일정하게 사용되는가?
  - 선행/후행 공백이나 탭이 없는가?

- 문자 인코딩 오류가 없는가?
- 대문자와 구두점이 일관되게 사용되며, 표준 규칙을 따르는가?
- 데이터 공백 값은 데이터 필드의 형식 및 내용 정의에 따라 적절한 코드를 사용하여 표시되는가?
- 문서화 및 처리 단계에서 값 및 계산 방법을 설명하는 정보가 포함되어 있는가?
- 데이터에 개인정보, 민감정보가 포함되어 있지 않은가?

### 【 데이터 진본성 】

- ☐ 디지털 정보의 경우, 복제 및 변경이 용이하므로 데이터의 신뢰성 입증과 무단 접근 방지와 관련한 대안이 필요함

### 【 품질 관리 사항 】

- ☐ 한국지질자원연구원은 연구 데이터셋의 최종 승인 완료 시점에 등록 API를 통해 DOI 또는 IGSN을 발급하여 식별 및 활용 추적, 관리를 진행함
  - 시료 데이터셋은 IGSN을 발급함
  - 시료 데이터셋이 아닌 그 외의 유형에 대해서는 DOI를 발급함
- ☐ 한국지질자원연구원은 2015년에 아시아 최초로 IGSN 등록기관으로 가입되었으며<sup>6)</sup> 암석, 시추코어 등 지질시료에 국제 공인된 유일식별번호를 부여할 수 있어 연구 자료 분석에 있어 보다 높은 신뢰성의 연구데이터를 보장할 수 있음

### 【 품질 관리 사항 문서화 시 기재 사항 】<sup>7)</sup>

- ☐ 품질 관리 사항을 문서화 시에는 다음과 같은 항목을 기재해야 함
  - 연구데이터 품질 평가 결과
  - 품질 평가를 수행한 담당자, 기술 요구사항 및 교육 기록
  - 품질 평가 시 사용된 방법
  - 사용한 연구데이터 품질 지표
  - 연구데이터 검증 절차
  - 연구데이터 스크리닝(screening) 방법

### 【 FAIR 원칙 】

- ☐ 연구데이터는 FAIR 원칙(Findable, Accessible, Interoperable, Reusable)을 준용함

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

7) USGS(n.d.). 데이터 품질 문서화: 고려사항.

[https://www.usgs.gov/data-management/documenting-data-quality-considerations?qt-science\\_support\\_page\\_related\\_con=0#qt-science\\_support\\_page\\_related\\_con](https://www.usgs.gov/data-management/documenting-data-quality-considerations?qt-science_support_page_related_con=0#qt-science_support_page_related_con)

□ FAIR 원칙은 최대한의 사용 및 재사용을 가능하게 하는 방식으로 데이터 공유에 대한 유용한 프레임워크를 제공함

□ Findability(검색가능성)

- 데이터에는 충분한 메타데이터와 다른 사람들이 쉽게 발견할 수 있는 고유하고 영구적인 식별자가 할당되어야 함
- 영구 식별자(예: DOI)를 할당하고, 데이터를 기술(description)할 충분한 메타데이터를 보유하고, 국제 검색 포털을 통해 데이터를 찾을 수 있는지 확인하는 작업이 포함됨

□ Accessibility(접근가능성)

- 데이터는 표준화된 통신 프로토콜을 통해 인간과 기계 모두에게 검색될 수 있으며, 필요한 경우 인증 및 승인이 가능함
- 데이터는 개인 정보 보호 문제, 국가 안보 또는 상업적 이익으로 인해 민감할 수 있음
- 데이터를 개방 할 수 없는 경우 접근 및 재사용을 규율하는 조건에 대한 명확성과 투명성이 있어야 함

□ Interoperability(상호운용성)

- 관련 데이터 및 메타데이터는 지식 표현을 위해 공식적이고, 접근하기 쉽고, 공유되고, 광범위하게 적용 가능한 언어를 사용함
- 데이터 및 메타데이터에서 커뮤니티에서 허용하는 언어, 형식 및 어휘를 사용하는 것이 포함됨
- 메타데이터는 식별자를 통해 다른 데이터, 메타데이터 및 정보와의 관계를 참조하고 설명해야 함

□ Reusability(재사용 가능성)

- 관련 메타데이터는 풍부하고 정확한 정보를 제공하며 데이터에는 명확한 사용 라이선스 및 자세한 출처 정보가 함께 제공됨
- 재사용 가능한 데이터의 메타데이터는 초기에 입력된 풍부하고 정확한 상태를 유지해야 함
- 하나의 특정 간행물에서 발견을 설명하기 위한 목적으로 재사용 가능성이 감소되어서는 안 됨
- 명확한 기계 판독 가능한 라이선스와 데이터 생성 방법에 대한 출처 정보가 필요함
- 분야별 데이터 및 메타데이터 표준을 사용하여 재사용을 허용하는 풍부한 맥락 정보를 제공해야 함
- 연구데이터와 메타데이터를 검색, 접근, 상호 운용, 재사용할 수 있는 원칙을 의미함

## 【 FAIR 데이터 체크리스트 】<sup>8)</sup>

### ☐ Findable(검색 가능)

- 데이터 검색 가능성(DOI 또는 Handle과 같은)을 지정하는 것은 데이터를 설명하는 풍부한 메타데이터를 가지며, 분야별 또는 국내외의 검색 포털을 통해 데이터를 찾을 수 있도록 하는 것을 포함함
- 다음 <표 4>는 Findable 체크리스트를 나타낸 것임

<표 4> Findable 체크리스트

체크리스트	내용	점수
데이터셋에 할당된 식별자가 있는가?	• 식별자 없음	0
	• 로컬 식별자	1
	• 웹 주소(URL)	2
	• 인용할 수 있으며 지속적인 국제적 고유 식별자(예: DOI, PURL, ARK 또는 Handle)	3
데이터를 설명하는 모든 메타데이터 레코드/파일에 데이터셋 식별자가 포함되어 있는가?	• 아니오	0
	• 예	1
메타데이터로 데이터를 어떻게 설명하는가?	• 데이터가 기술되어 있지 않음	0
	• 약칭 및 설명	1
	• 포괄적이지만 텍스트 기반, 비표준 형식	2
	• 공인된 공식 머신러닝 메타데이터 스키마를 종합적으로 활용(제안 참조)	3
메타데이터 레코드가 있는 리포지터리 또는 레지스트리 유형은 무엇인가?	• 데이터가 리포지터리에 설명되지 않음	0
	• 로컬 기관 리포지터리	1
	• 도메인별 리포지터리	1
	• 제너럴리스트 공용 리포지터리	1
	• 데이터가 한 곳에 있지만 여러 레지스트리를 통해 검색 가능	2

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

□ Accessible(접근 가능)

- 데이터에 접근할 수 있도록 하는 것은 표준화된 프로토콜을 사용하여 데이터를 개방하는 것을 포함할 수 있음
- 개인정보보호, 국가 안보 또는 상업적 이익과 같이 데이터를 개방할 수 없는 타당한 이유가 있을 경우, 접근 및 재사용을 제한할 수 있고, 이를 위해서는 접근 제한에 대한 명확성과 투명성이 있어야 함
- 다음 <표 5>는 Accessible 체크리스트를 나타낸 것임

<표 5> Accessible 체크리스트

체크리스트	내용	점수
데이터에 얼마나 쉽게 접근할 수 있는가?	• 데이터 또는 메타데이터에 접근할 수 없음	0
	• 메타데이터만 접근 가능	1
	• 지정되지 않은 조건부 접근(예: 데이터 관리자 접근)	2
	• 지정된 날짜 이후 접근 금지	3
	• 식별되지 않은/수정된 데이터의 하위 집합은 공개적으로 접근 가능	4
	• 민감한 데이터에 대한 윤리 승인과 같이 명시된 조건을 충족하는 사람이 완전히 접근 가능	5
	• 공개적으로 접근 가능	5
접근이 승인되면 특수 프로토콜이나 도구를 필요로 하지 않고 온라인으로 데이터를 이용할 수 있는가?	• 데이터 접근 금지	0
	• 개별약정에 따라 이용	1
	• 온라인에서 파일 다운로드	2
	• 비표준 웹 서비스(예: OpenAPI, Swagger, 비공식 API)	3
	• 표준 웹 서비스 API(예: OGC)	4
데이터를 더 이상 사용할 수 없는 경우에도 메타데이터 레코드를 사용할 수 있는가?	• 확실하지 않음	0
	• 아니오	0
	• 예	1

□ Interoperable(상호운용가능)

- 상호운용성을 위해서는 데이터가 커뮤니티의 합의 형식, 언어 및 어휘를 사용해야 함
- 메타데이터는 커뮤니티가 합의한 표준과 어휘를 사용해야 하며 식별자를 사용하여 관련 정보에 대한 링크를 포함해야 함
- 다음 <표 6>은 Interoperable 체크리스트를 나타낸 것임

<표 6> Interoperable 체크리스트

체크리스트	내용	점수
데이터는 어떤 형식으로 제공되는가?	• 대부분 독점 형식	0
	• 구조화된 개방형 표준, 비기계적 형식	1
	• 구조화된 개방형 표준, 기계 제어 형식	2
데이터 요소를 정의하는 데 사용되는 어휘/온톨로지/태깅 스키마의 유형을 가장 잘 설명하는 것은 무엇인가?	• 데이터 요소가 설명되지 않음	0
	• 데이터 요소 설명에 표준이 적용되지 않음	1
	• 국제 식별자가 없는 표준화된 어휘/온톨로지/언어	2
	• 설명에 연결되는 해석 가능한 국제 식별자를 사용하여 개방적이고 보편적인 표준화	3
메타데이터를 다른 데이터 및 메타데이터와 어떻게 연결하는가?	• 다른 메타데이터에 대한 링크 없음	0
	• 메타데이터 레코드에는 관련 메타데이터, 데이터 및 정의에 대한 URI 링크가 포함됨	1
	• 메타데이터는 기계 판독 가능한 형식(예: RDF(Resource Description Framework)과 같은 링크된 데이터 형식으로 표시됨	2

□ Reusable(재사용 가능)

- 재사용 가능한 데이터는 입력된 초기 메타데이터의 풍부함과 정확함을 유지해야 함
- 정확한 해석과 재사용이 가능한 풍부한 상황별 정보를 제공하기 위해 분야별 데이터와 메타데이터 표준을 사용해야 함
- 다음 <표 7>은 Reusable 체크리스트를 나타낸 것임

<표 7> Reusable 체크리스트

체크리스트	내용	점수
다음 중 데이터에 첨부된 라이선스/사용 권한을 가장 잘 설명한 것은 무엇인가?	• 라이선스 없음	0
	• 비표준 텍스트 기반 라이선스	1
	• 비표준 머신러닝 라이선스 (데이터를 어떤 조건에서 재사용할 수 있는지를 나타내는 라이선스)	2
	• 표준 텍스트 기반 라이선스	2
	• 표준 기계판독형 라이선스 (예: Creative Commons)	3
데이터 재사용을 촉진하기 위해 얼마나 많은 환경 정보가 수집되었는가?	• 환경 정보가 기록되지 않음	0
	• 일부(부분적) 기록	1
	• 텍스트 형식으로 전체 기록	2
	• 기계 판독 가능한 형식으로 완전 기록	3



## 【 연구데이터 보존 】

- ☐ 한국지질자원연구원은 연구데이터의 보존을 위한 지침을 작성하여 이를 준수하고 있으며, 연구데이터의 보존에 대한 자세한 사항은 해당 지침에 기술되어 있음
  - 연구데이터 보존 지침에는 연구데이터 보존 개념, 보존 데이터와 리포지터리의 선택 및 평가 등의 내용이 포함됨
- ☐ 연구데이터 관리 지침에서는 연구데이터의 보존을 위한 메타데이터 기술을 중심으로 한 내용이 서술됨

## 【 연구데이터의 보존과 기술(Description) 】

- ☐ 연구데이터 보존 단계에서는 다음과 같은 활동이 요구됨
  - 데이터의 관리 특성을 장기간 보존하고 보유하기 위한 조치
  - 데이터를 장기적으로 보존할 계획 수립
  - 보존할 데이터, 보존할 위치 및 데이터와 함께 필요한 도큐멘테이션 결정
  - 보존을 위한 메타데이터 및 도큐멘테이션 생성
  - 데이터 구성 및 저장

## 【 메타데이터 의미 】

- ☐ 메타데이터는 데이터의 설명에 필요한 데이터의 제목, 데이터의 생산자, 데이터의 생산 장비 및 방법, 데이터의 내용, 획득지역(위치좌표) 및 시기, 데이터의 포맷, 데이터의 품질 등으로 구성되어 데이터를 설명하기 위해 사용되는 데이터를 의미함

## 【 메타데이터 기술 시 고려 사항 】

- ☐ 한국지질자원연구원에서 정의한 메타데이터 스키마를 준용함
- ☐ 연구데이터의 장기간 보존 및 활용을 위해 풍부하게 기술함
- ☐ 메타데이터 레코드를 어떻게 생성하는지 확인함
- ☐ 어떤 메타데이터 표준을 사용하는지 확인함
- ☐ 어떤 도구를 사용하는지 확인함

☐ 프로젝트가 시작될 때 레코드를 생성하고 연구를 진행하면서 업데이트하는지 확인함

☐ 메타데이터를 어디에 기탁할지 고려함

☐ 메타데이터 표준 및 리포지토리를 결정할 때 커뮤니티의 표준을 고려함

**【 메타데이터 기술 요소】<sup>9)</sup>**

- ☐ 수집 및 기탁 시 사용되는 메타데이터는 DC(Dublin Core)의 형식을 사용함
- DC 메타데이터 요소 15개에 대한 사항은 연구데이터 수집 지침을 따름

---

9) 국립중앙도서관. DCMI 메타데이터 용어. <https://librarian.nl.go.kr/LI/contents/L10202000000.do>

## 【 의 미 】

- ☐ 연구데이터의 재사용을 위해서는 데이터를 검색하고 접근할 수 있어야 함
- ☐ 연구데이터 리포지터리는 이용자가 연구데이터 검색 및 접근을 쉽게 할 수 있도록 기능을 제공해야 함
- ☐ 리포지터리를 사용하면 시간이 지남에 따라 디지털 객체를 재사용할 수 있으므로 이해 및 사용을 지원하는 데 적합한 정보를 사용할 수 있음
- ☐ 검색 가능한 데이터에 제한 없이 접근할 수 있는 것은 아니므로 기밀성 유지, 재사용 허가, 접근 제한, 라이선스 등이 존재할 수 있음

## 【 연구데이터 공개 및 공유 】

- ☐ 연구자(데이터 생산자)가 연구데이터를 제출 시 비공개, 내부공개, 대외공개를 선택할 수 있으며, 엠바고를 통해 공개 시한을 지정할 수 있음
- ☐ 연구자는 연구기밀이 보호되고 관련 규정이 허용되는 범위 내에서 가능한 한 많은 사람들이 연구데이터를 이용할 수 있도록 공개 또는 공유해야 함
- ☐ 연구데이터를 공개 또는 공유할 때는 다른 연구자가 손쉽게 재사용할 수 있도록 공통 표준에 맞추어 알아보기 쉽게 정리하고 적절한 조건의 라이선스와 함께 제공해야 함
- ☐ 연구자는 개인정보가 포함되어 있는 민감 데이터를 보호하고 관리하기 위해 비식별화 형태로 변형한 이후 공개 또는 공유해야 함

【 연구데이터 공유의 중요성 및 필요성 】<sup>10)</sup>

- ☐ 다양한 측정 및 실험 장비로부터 생산되는 방대한 데이터를 통해 과학적 발견을 시도하는 데이터 집종형 연구로 전환함에 따라, 연구데이터의 재활용과 관리의 중요성이 증대됨
- ☐ 오픈 사이언스, 오픈 액세스 운동 등과 더불어, 공공 자금에서 출자한 연구 과제를

10) DataON. 연구데이터 관리활용 가이드라인. 출처: [https://dataon.kisti.re.kr/data\\_mgnt\\_guideline\\_08.do](https://dataon.kisti.re.kr/data_mgnt_guideline_08.do)

통해 생산된 데이터와 출판물은 공용 리포지터리에서 게시하는 형태로 전환 중

- ☐ 연구데이터 공유는 다른 연구자 또는 기관에 데이터 활용도를 높임
- ☐ 데이터 인용(Data Citation)등을 통해 연구자 개인 또는 연구 기관의 평판을 높이고, 데이터 검증 등을 통한 더 나은 연구로 발전할 수 있는 기회가 됨
- ☐ 데이터의 공유는 데이터의 중복 생산, 중복 게재 비용을 줄일 수 있으며, 이를 통해 미래의 연구에 더욱 집중할 수 있음
- ☐ 사생활 보호, 연구 윤리 등의 민감한 사항들을 잘 고려하여 전 처리된 대부분의 데이터 공유가 가능함
- ☐ 연구개발기관이 DMP 등을 통해 연구데이터 공개를 추진하면, 데이터 수집, 저장, 메타데이터 작성 표준화는 연구데이터 공유를 실현하기 위한 연구데이터 관리의 최선의 방안임
- ☐ 연구데이터의 공유 및 관리를 위한 데이터 리포지터리에 대한 자세한 내용은 한국지질자원연구원 연구데이터 보존 지침에 명시되어 있음
  - 연구데이터 보존 지침에는 데이터 리포지터리의 정의, 선택 시 고려 사항, 리포지터리 예시 등이 포함됨

#### 【 FAIR 원칙: Reusable 】

- ☐ FAIR 원칙은 연구데이터를 가능한 한 많이 재사용하고 공유하도록 생각할 수 있는 유용한 프레임워크를 제공하며, Reusable(재사용 가능) 항목은 연구데이터의 공유 및 활용에 대한 상세한 체크리스트를 제시함
- ☐ 아래는 Reusable 항목의 자세한 내용임
  - 연구데이터 관련 메타데이터는 풍부하고 정확한 정보를 제공하며 연구데이터에 명확한 사용 라이선스 및 자세한 출처 정보가 함께 제공되어야 함
    - 재사용 가능한 연구데이터는 요약물 하지 않아야 하며, 원시 데이터의 풍부함을 유지해야 함
    - 명확한 기계 판독 가능한 라이선스와 데이터 생성 방법에 대한 출처 정보가 필요함
    - 분야별 데이터 및 메타데이터 표준을 사용하여 재사용을 허용하는 풍부한 맥락 정보를 제공해야 함
  - 재사용이 가능하도록 정확하고 관련성이 높은 여러 속성이 제공되어야 함
  - 연구데이터와 메타데이터 모두 기계가독형이고, 커뮤니티 표준을 준수하여야 함

- 메타데이터에는 연구데이터를 재사용할 수 있는 라이선스에 대한 정보가 포함되어야 함
- 메타데이터는 표준 재사용 라이선스를 참조하여야 하며, 이것은 시스템이 이해할 수 있는 재사용 라이선스를 의미함
- 메타데이터에는 커뮤니티별 표준에 따른 출처 정보와 커뮤니티 간 언어에 따른 출처 정보가 포함되어야 하며, 기계판독이 가능한 형식이어야 함

## 【 엠바고(Embargo) 】<sup>11)</sup>

- ☐ 연구자가 지정한 시간이 경과할 때까지 제목, 작성자, 메타데이터, 초록 등의 데이터셋에 관련한 설명은 열람할 수 있으나, 데이터셋에 대한 공개적인 액세스는 불가능한 기간을 의미하며, 엠바고 기간이 끝날 시 공개 또는 중재(mediated: 신청서 승인 이후 이용 가능한 데이터) 액세스를 통해 데이터셋을 사용할 수 있음<sup>12)</sup>
- ☐ 데이터 리포지터리는 연구자에 의해 설정된 엠바고 정보를 관리할 의무가 있음
- ☐ 데이터셋을 엠바고 상태로 기탁할 경우, 엠바고 제공일 및 종료일이 제시되어야 함. 데이터 리포지터리는 엠바고 기간이 종료될 때까지 데이터셋에 대한 액세스를 제한하며, 이용자는 엠바고 종료일 이후 데이터셋의 사용이 가능함
- ☐ 데이터셋에 엠바고를 적용할 경우 이용자의 데이터셋의 액세스가 제한되므로, 연구자에게 엠바고 적용에 대한 사례나 근거가 요청될 수 있음
- ☐ 데이터셋에 엠바고를 적용하는 상황으로는 다음과 같은 예가 포함됨
  - 데이터셋에 액세스할 수 있도록 하기 전에, 데이터셋에 기반한 연구를 출판할 경우
  - 특허 출원 등 공개 전 연구데이터로 인한 상업적 이익을 마무리해야하는 경우
  - 데이터셋이 데이터 민감성과 관련된 경우

## 【 연구데이터 인용 】

- ☐ 연구데이터 인용이란, 연구자가 연구 출판물에서 학술 기사, 보고서와 같은 연구 산출물에 대한 서지 참조를 제공하는 것과 같은 방식으로 데이터에 대한 참조를 제공하는 것을 의미함

11) 국가과학기술연구회(2018). 정책과제-2018-09\_출연(연) 연구데이터 관리·활용 방안 연구.pdf. 출처: <https://www.nst.re.kr/www/selectBbsNttView.do;jsessionid=0C1DA6F31EBF881198E95FF8105B74AB?key=64&bbsNo=34&nttNo=2205&searchCtgry=&searchCnd=all&searchKrwD=&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>

- ☐ 연구데이터 인용은 데이터를 주요 연구 결과로 인식하는 핵심 사례 중 하나로 인식되고 있음

### 【 연구데이터 인용의 필요성 】

- ☐ 인용 데이터셋은 연구데이터 생성자 및 연구데이터 관리자에게 보상(acknowledgement)을 제공함
- ☐ 연구자에게 적절한 크레딧(credit)을 제공하고 학문적 노력의 보상으로 작용하며 장기간 연구데이터를 관리하는 연구데이터 관리자와 리포지터리에 크레딧을 제공함
- ☐ 연구데이터 인용은 연구데이터셋에 대한 생성자 및 관리자에게 책임을 부여하고, 데이터셋이 적절하게 인용되면 표절에 대한 위험을 줄일 수 있음
- ☐ 연구데이터 인용을 통해 연구 결과의 데이터 검색, 재사용 및 검증이 가능하며, 이러한 기능은 연구데이터셋의 검색 가능성과 재사용성을 향상시킴
- ☐ 출판된 연구데이터를 이용할 때, 연구데이터 기탁자가 지정한 저작권 및 라이선스 조건을 확인하고 준수하여야 함
- ☐ 연구데이터 인용은 과학 커뮤니티에 연구데이터 생산자에 대한 인식과 보상의 공식적인 체계를 형성함
- ☐ 연구데이터셋을 인용하여야 해당 연구데이터셋을 추적 및 계산할 수 있으며, 이를 통해 영향력 지수(Impact factor)와 같은 연구데이터셋의 영향력 평가가 가능함
- ☐ 연구데이터 인용은 출판물에서 연구데이터 생산의 투명성을 증가시킬 뿐만 아니라, 고품질의 연구데이터셋 생산을 장려함
- ☐ 연구데이터 인용이 일상화되면 연구데이터를 최고 수준의 연구 결과로 재현 가능함
- ☐ 학술 논문과 비슷한 방식으로, 영향력 지수(Impact factor)를 측정하기 위하여 필요함
- ☐ 인용 내에서 영구 식별자를 사용하는 것은 모든 연구데이터 인용이 정확하게 추적되고 계산되도록 하는 핵심 요소임

- ☐ 인용 지수는 연구데이터의 재사용을 측정하며, 이러한 측정기준(metric)은 일반적으로 성과 평가 및 보고에 사용될 수 있음

### 【 연구데이터 인용의 구성요소 】

- ☐ 연구데이터 인용에는 다음과 같은 구성요소들이 있음
  - 저자
  - 제목
  - 공개날짜
  - 버전 또는 에디션 번호
  - 배포자
  - 영구 식별자
  - 액세스 날짜 및 시간

### 【 연구데이터 식별자】

- ☐ 데이터 식별자는 모든 컴퓨터 기반 시스템에서 필수적이며, 학술 기록을 인용할 수 있도록 함
- ☐ 영구 식별자란 모든 종류의 객체를 고유하게 식별하는 식별자이며, 지속적인 관리 및 최신 상태로의 유지가 필요함
- ☐ 연구데이터 인용에 주로 사용되는 영구 식별자로, DOI가 있음
- ☐ DOI의 이점<sup>13)</sup>
  - 품질과 정확도에 있어 높은 수준의 신뢰성을 보장함
  - IDF(International DOI Federation) 및 등록 기관 인프라에서 지원됨
- ☐ 데이터 인용
  - DOI에는 명확한 데이터 인용을 생성하는 메타데이터 요소가 필요함
    - 예) Hanigan, Ivan (2012): Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index. The Australian National University Australian Data Archive. DOI (doi. org/10.4225/13/50bbfd7e6727a)
  - 영구성
    - DOI는 데이터셋이 잘 관리되고 장기적으로 사용할 수 있음을 나타냄
  - 접근 가능성
    - DOI는 연구데이터에 대한 온라인 접근성을 용이하게 함

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

### 【 연구데이터 인용 방법】

- ☐ 출판된 연구데이터는 다른 학술적 산출물과 동일한 방식으로 인용해야 함
- ☐ 연구데이터 플랫폼에 등록 및 수집된 데이터를 출판 및 인용 시 플랫폼 데이터 인용을 명기해야 함
- ☐ 학술 기사 인용 스타일과 형식이 다양한 것처럼 데이터의 인용 스타일도 다양함
- ☐ 데이터 인용은 일반적으로 저자, 공개일, 데이터 제목, 버전 또는 에디션 번호, 아카이브 또는 배포자, 영구 식별자, 액세스 날짜 및 시간 등의 요소로 구성됨
- ☐ 표준 인용은 다음의 요소를 포함하고 있음(\*은 선택임)

저자. (출판년도). 제목. 버전\*. 출판사. DOI\*

〈그림 2〉 연구데이터 표준 인용 형식

- ☐ 지오빅데이터 오픈플랫폼의 데이터 인용 방법은 다음과 같음<sup>14)</sup>
  - 연구데이터를 제공받은 자가 그 활용 결과물을 논문, 보고서, 웹사이트 등을 통하여 공표할 경우 결과물에 데이터 인용 문구를 포함하여야 함
  - 데이터 인용문구 형식
    - 저자. (데이터 생산년도). 데이터셋 명. 데이터발행기관. DOI
    - 예: 홍길동. (2020.11.01.). 적외선 분광 분석데이터. 미상. <http://doi.org/----->
- ☐ 지오빅데이터 오픈플랫폼의 데이터 인용 시 이용조건에 대한 사항은 연구데이터 관리 규정(제6장 연구데이터의 개방 및 이용조건)에 명시되어 있으며, 내용은 다음과 같음
  - 연구데이터를 이용하는 사람은 연구데이터를 생산한 사람의 공헌도를 알려야 하며, 연구데이터 라이선스에서 별도로 지정하지 않는 경우, 인용 문구를 사용하여 공식적으로 연구데이터 이용을 명시해야 함
  - 연구데이터를 이용하는 사람은 이용하고자 하는 연구데이터에 라이선스가 적용된 경우 반드시 해당 라이선스 이용조건을 확인하고, 저작자, 출처 등을 명시하는 등 이용조건을 준수해야 함

14) KIGAM 연구보고서: 「지오빅데이터 오픈플랫폼 개발을 통한 국가지질자원데이터센터(NGDC) 기반 구축」의 부록 4: 연구데이터 개방 및 제공 지침



## 【 연구데이터 출판 】

### ☐ 출판 절차

- 데이터 출판은 다른 연구자들이 (재)활용할 수 있도록 연구데이터를 공개하는 행위로, 연구기관 등의 데이터 리포지토리를 통해 온라인으로 접근 가능한 형태로 출판될 수 있음
- 관련 주제 분야 혹은 데이터 전문가의 피어 리뷰 등 별도의 품질 점검 절차를 도입할 수 있음

### ☐ 데이터 식별자 부여

- 데이터셋의 식별 및 활용 추적을 위해서 공개된 연구데이터에 대해서는 DOI 및 IGSN 등과 같은 식별자가 부여됨

### ☐ 라이선스 적용

- 지오빅데이터 오픈플랫폼은 연구데이터 공개와 저작권 확보를 위해 크리에이티브 커먼즈 라이선스 기준과 지원 서비스를 제공하고 있으며, 이용자가 각 연구데이터에 대해 적절한 라이선스를 선택하여 적용할 수 있음
- 기타 내용은 한국지질자원연구원 윤리·저작권·라이선스 지침을 따름

버전	일자	내용
0.1	2023. 03. 20.	문서 Outline 작성
0.6	2023. 04. 28.	분야별 초안 작성
0.8	2023. 05. 08.	지침 기관 검토
1.0	2023. 05. 19.	검토 의견 반영 및 보완