

# Geoscience 연구데이터 관리를 위한 기능별 세부요소 및 중요도에 관한 연구\*

## A Study on Functional Details and Importance of Geoscience Research Data Management

김 주 섭 (Juseop Kim)\*\*

김 선 태 (Suntae Kim)\*\*\*

최 상 기 (Sangki Choi)\*\*\*\*

### 목 차

- |                                  |   |
|----------------------------------|---|
| 1. 서론                            | 4. Geoscience 연구데이터 관리의 기능별<br>세부요소 구성과 중요도 |
| 2. 이론적 배경                        | 5. 결론                                       |
| 3. 연구데이터 관리의 기능별 세부요소<br>조사 및 분석 |   |

### 초 록

본 연구는 국내에서 Geoscience 분야 연구데이터 관리 시스템 개발에 적용할 수 있는 RDM 기능별 세부요소의 도출을 목적으로 하였다. 연구 목적 달성을 위해 RDM 서비스와 관련된 8개의 기관을 분석하였다. 분석한 결과, Geoscience RDM 기능의 80개 세부요소가 도출되었으며, 도출된 세부요소를 검증하기 위하여 국내 전문가를 대상으로 설문조사를 수행하였다. 조사 결과, Geoscience 분야를 위한 80개의 RDM 기능별 세부요소가 중요도 순으로 제시되었다. 제시된 요소는 국내의 연구 기관 또는 대학도서관에서 Geoscience 분야의 RDM 서비스 구축 및 운영 시 기능별 세부요소로 제시될 수 있으며 순위에 따라 중요도를 판단할 수 있는 참고자료가 될 수 있을 것이다.

### ABSTRACT

The purpose of this study is to derive the detailed elements of each RDM function that can be applied to the development of research data management system in Geoscience field in Korea. Eight institutions related to RDM services were analyzed to achieve the research purpose. As a result of the analysis, 80 detailed elements of Geoscience RDM function were derived, and a survey was conducted to domestic experts to verify the derived details. As a result, 80 RDM functional details for Geoscience are presented in order of importance. The elements presented can be presented as functional details in the establishment and operation of RDM services in the field of Geoscience in research institutes or university libraries in Korea.

키워드: 연구데이터, 연구데이터 관리, 오픈 사이언스, 지오사이언스, 큐레이션

Research Data, Research Data Management (RDM), Open Science, Geoscience, Curation

\* 본 연구는 전북대학교 대학원 박사학위논문 중 일부를 발췌하여 요약·수정한 것임.

\*\* 전북대학교 문헌정보학과 시간강사(jusup.kim@gmail.com / ISNI 0000 0004 7492 1806) (제1저자)

\*\*\* 전북대학교 문헌정보학과 조교수(kim.suntae@jbnu.ac.kr / ISNI 0000 0004 6492 6355) (교신저자)

\*\*\*\* 전북대학교 문헌정보학과 교수, 문화융복합아이빙연구소 연구원  
(choisk@jbnu.ac.kr / ISNI 0000 0004 6468 1351) (공동저자)

논문접수일자: 2020년 1월 22일 최초심사일자: 2020년 2월 10일 게재확정일자: 2020년 2월 19일

한국문헌정보학회지, 54(1): 411-440, 2020. <http://dx.doi.org/10.4275/KSLIS.2020.54.1.411>

## 1. 서론

### 1.1 연구의 필요성 및 목적

최근 국내외에서 공공기금으로 연구된 연구 결과물이 국가 자산 및 기관 자산으로 재사용될 수 있도록 체계적으로 관리 및 보존되어야 한다는 인식이 확산하고 있다. 여기에서 연구 결과물은 학술지뿐만 아니라 연구 과정에서 생산된 데이터까지 포함되며, 이러한 데이터를 연구데이터라고 명명하여 국외에서는 이미 체계적으로 관리할 수 있는 서비스를 계획하거나 구현하는 추세다. 특히, 연구데이터는 연구 과정에서 생성되는 모든 데이터를 말하며 연구 패러다임이 데이터-기반 연구로 변화함에 따라 더욱 중요하게 인식하고 있다.

이러한 데이터 중심 연구의 배경으로 자주 등장하는 오픈 사이언스(Open Science)는 연구를 가속하기 위한 수단으로서 공개적으로 자금을 지원하는 연구 결과(출판물 및 연구데이터)의 주요 산출물을 디지털 형식으로 접근할 수 있도록 만드는 연구자, 정부, 연구비 지원 기관 또는 과학 커뮤니티 자체의 노력을 의미한다(OECD 2015, 7). 오픈액세스(Open Access), 오픈데이터(Open Data), 오픈소스(Open Source) 등으로 구체화하는 오픈 사이언스 운동에서도 데이터가 그 핵심에 있다(Kraker et al. 2011). 이 중에서 오픈데이터는 두 가지 측면에서 큰 의미가 있는데 하나는 과학의 활성화 측면에서의 의미이며, 또 하나는 '국가 데이터 개방과 재 활용' 측면에서의 의미이다. 학술 연구 영역에서는 이러한 데이터를 '연구데이터'라 부르며, 국가의 세금으로 수행된 연구 산출물에 대해

법적, 제도적, 기술적 장벽 없이 접근과 활용이 가능하도록 하자는 오픈 사이언스 운동은 주요 선진국에서 다양한 형태의 활동을 끌어내고 있다(김주섭, 김선태, 최상기 2019).

국외의 경우 미국, 영국, 호주 등 연방기관, 연구회를 중심으로 국가 차원에서 연구데이터 공유·활용 정책을 시행하고 인프라를 구축하여 운영하는 추세다(이상환 2019). 여기에 연구데이터가 국가와 기관의 자산이라는 인식이 확산하면서 체계적인 데이터 관리와 재사용을 보장하기 위해 데이터 관리 계획(Data Management Plan, 이하 DMP)이 요구되기 시작하면서 연구 제안 시 연구비 지원 기관에 연구 계획서와 함께 DMP 제출을 요구받게 되었다. 더불어 연구기관들은 오픈데이터, DMP 요구 사항 및 잠재적인 '빅데이터(big data)' 연구 기회에 대한 중요성이 커짐에 따라 연구 프로세스의 많은 지점에서 요구 사항을 충족시키기 위해 더 광범위하고 세련된 데이터 서비스를 제공해야 한다는 새로운 요구를 인식하고 있다(Fearon et al. 2013). 즉, 대학도서관 및 연구기관에서 연구자를 지원하기 위한 연구데이터의 탐색, 공유, 접근 및 보존 등을 지원할 수 있는 연구데이터 관리 및 서비스(Research Data Management Service, 이하 RDMS)가 필요하다.

아직 국내에서는 전반적인 연구데이터 관리에 관한 인식이 부족하며 특히 특정 학문 분야의 연구데이터를 관리한다는 건 더욱 생소한 실정이다. 연구데이터의 특성상 해당 연구 분야에 의존적인 성격으로 범용적인 연구데이터 관리보다는 특정 학문 분야를 위한 연구데이터 관리가 필수적이다.

국외의 경우, 바이오, 물리학, Geoscience 분

야 그리고 천문학 등의 학문 분야에서 DMP, 리포지토리 등과 같은 연구데이터를 관리하기 위한 움직임이 타 분야보다 일찍 시작되었다(신은정 2018). 특히, Geoscience 분야 데이터는 타 학문보다 많은 시간과 노력 그리고 값비싼 장비를 통해 실시간으로 생산되고 있어 해당 데이터에 대한 관리가 중요하다고 할 수 있다. 또한 Geoscience 분야는 자연환경의 지속적인 변화를 조사하기 위하여 데이터 취득을 지속해서 수행해야 하는데 여기에는 인력, 시간 그리고 많은 예산이 필요하며 미래예측에 필요한 과거 데이터의 보존 및 활용이 중요하다고 할 수 있다(한종규 2019). 이에 본 연구는 Geoscience 분야의 연구데이터 관리(Research Data Management, 이하 RDM) 시스템 개발에 적용할 수 있는 RDM 기능별 세부요소의 도출을 목적으로 하였다.

### 1.2 연구 방법 및 절차

본 연구의 목적을 달성하기 위해 사용된 연구 방법 및 절차는 다음과 같다. 다음 <그림 1>은 본 연구의 내용 및 절차를 도식화한 것이다.

1단계	연구 방향 및 범위 설정
2단계	RDM 관련 해외 선진 기관 분석 - DCC 등
3단계	RDM 기능별 세부요소 도출
4단계	RDM 기능별 세부요소 중요도 파악 - 설문조사
5단계	Geoscience RDM 기능별 세부요소 제시

<그림 1> 단계별 연구 내용 및 절차

첫 번째 단계에서는, 전반적인 연구의 방향

과 범위를 설정하고 두 번째 단계에서는 DCC (Digital Curation Centre) 등 RDM(Research Data Management, 이하 RDM)과 관련한 8개의 국외 선진 기관을 분석하였다. 세 번째 단계에서는 국외 선진 기관과 Geoscience 분야의 전문가를 통해 도출한 RDM의 기능별 세부요소를 정리하였다. 네 번째 단계에서는 도출된 RDM 기능별 세부요소의 중요도를 파악하기 위하여 국내의 정보 유통 분야의 전문가들을 대상으로 설문조사를 실시하였다. 마지막으로 설문조사를 통해 나타난 세부요소의 우선순위에 따라 Geoscience 분야 RDM 기능별 세부요소를 제시하였다.

본 연구의 목적을 수행하기 위한 구체적인 연구 질문은 다음과 같다.

- 연구 질문: 국내에 적용 가능한 Geoscience 분야 연구데이터 관리는 어떻게 구성될 수 있는가?
- 세부 연구 질문: 사례분석을 통해 Geoscience 연구데이터 관리를 위한 기능별 세부요소는 무엇인가?
- 세부 연구 질문: 도출된 세부요소에 대한 전문가 및 연구자의 중요도 평가는 어떠한가?

지금까지 국내에서는 국외 대학도서관이나 연구기관처럼 범용적인 RDM을 위한 서비스의 구축 및 운영 사례가 없는 실정이다. 이런 점에서 전적으로 국외사례를 중심으로 사례조사를 진행하였고 또한 모든 RDM 사례를 분석할 수 없어 8개 기관만 선택한 것이 본 연구의 제한점이라 할 수 있다.

## 2. 이론적 배경

### 2.1 연구데이터 관리(RDM)

새로운 유형의 과학으로 인해 발생하는 ‘데이터의 홍수’, 특정 분야의 연구 무결성에 대한 신뢰의 위기, 오픈 데이터 운동으로 인해 데이터를 더 잘 관리하고 더 많이 공유하는 것에 대한 관심이 높아졌다(Cox and Verbann 2018). 특히, 국가 세금으로 만들어진 연구 결과물에 대해 대학 및 기관에서는 연구 중 생산된 모든 데이터를 공개하려는 움직임이 한층 더 가속화되고 있다. 다음의 <표 1>은 문헌 등에서 확인한 RDM의 의미를 정리한 것이다.

RDM은 데이터 입력부터 연구 사이클, 귀중한 결과의 배포 및 보존에 이르기까지 데이터 구성과 관련이 있다. 이것은 연구 결과의 신뢰성 있는 검증을 보장하고 기존 정보를 바탕으로 새롭고 혁신적인 연구를 허용한다(Whyte and Tedds 2011).

또한, RDM은 프로젝트의 계획 단계에서 데이터의 장기 보존에 이르기까지 연구데이터를

처리하기 위해 연구 라이프 사이클(Research Life Cycle)동안 이루어진 모든 결정을 다룬다. 이러한 RDM을 미리 계획하면 데이터 품질을 보장하고 위험을 최소화하며 시간을 절약하고 법률, 윤리, 기관 및 펀더의 요구 사항을 준수할 수 있다(University College London 2019).

영국의 옥스퍼드 대학에서는 RDM에 대하여 “연구데이터 관리는 연구 프로젝트 중에 사용되거나 생성된 정보를 구성, 구조, 저장 및 관리하는 방법을 다루는 일반적인 용어”라고 기술하며, 여기에는 데이터 관리 계획, 프로젝트 기간 동안 정보를 처리하는 방법 그리고 연구가 종료된 후 데이터를 어떻게 관리해야 하는지를 포함하고 있다(Research Data Oxford 2019).

미국의 피츠버그 대학은 “RDM은 연구 프로젝트에서 수집 및 사용되는 데이터의 조직, 저장, 보존 및 공유를 설명하는 용어이며, 여기에는 일관된 파일 네이밍 규칙을 사용하여 연구 프로젝트 사이클 동안 연구데이터를 일상적으로 관리하는 작업이 포함된다. 또한 프로젝트가 완료된 후 데이터를 보존하고 공유하는 방법(예: 장기 보존 및 접근을 위해 데이터를 리

<표 1> RDM(Research Data Management)의 의미

출처	RDM의 의미
Whyte and Tedds	데이터 입력부터 연구 사이클, 귀중한 결과의 배포 및 보존에 이르기까지 데이터 구성과 관련이 있음
University College London	프로젝트의 계획 단계에서 데이터의 장기 보존에 이르기까지 연구데이터를 처리하기 위해 연구 라이프 사이클(Research Life Cycle)동안 이루어진 모든 결정을 다룬다
Research Data Oxford	연구데이터 관리는 연구 프로젝트 중에 사용되거나 생성된 정보를 구성, 구조, 저장 및 관리하는 방법을 다루는 일반적인 용어
University of Pittsburgh	연구 프로젝트에서 수집 및 사용되는 데이터의 조직, 저장, 보존 및 공유를 설명하는 용어
Eindhoven University of Technology	연구 활동을 효율적으로 수행하고 다른 사람들과 협력할 수 있도록 전체 연구 라이프 사이클 동안 연구데이터를 신중하게 처리하고 구성하는 것과 관련이 있음
Cox and Verbann	모든 연구 프로세스 내에서 데이터를 생성, 탐색, 구성, 저장, 공유 및 보존하는 것

포지토리에 저장)에 대한 결정도 포함된다.”라고 소개하며 다음과 같이 RDM의 중요성에 대하여 다음과 같이 제시하였다(University of Pittsburgh 2019).

- 데이터(특히 디지털 데이터)는 깨지기 쉽고 쉽게 손실됨
- 펀더와 출판사가 부과하는 연구데이터 요구 사항이 증가하고 있음
- 연구데이터 관리는 장기적으로 시간과 자원을 절약함
- 올바른 관리는 오류를 방지하고 분석 품질을 향상시킴
- 잘 관리되고 접근 가능한 데이터를 통해 다른 사람들이 결과를 검증하고 복제할 수 있음
- RDM은 연구데이터의 공유를 용이하게 하며, 공유될 때, 데이터는 다른 연구자들에 의해 귀중한 발견으로 이어질 수 있음

네덜란드 에인트호번 공과대학의 경우, “RDM은 연구 활동을 효율적으로 수행하고 다른 사람들과 협력할 수 있도록 전체 연구 라이프 사이클 동안 연구데이터를 신중하게 처리하고 구성하는 것과 관련이 있다”라며 다음과 같은 RDM의 목표를 제시하였다(Eindhoven University of Technology 2019).

- 연구데이터의 손실 방지
- 다른 사람들과의 연구데이터 공유
- 검색, 접근 그리고 재사용 가능한 연구데이터

Cox와 Verbann은 「연구데이터 관리 탐색」에서 “RDM은 모든 연구 프로세스 내에서 데

이터를 생성, 탐색, 구성, 저장, 공유 및 보존하는 것이며, 연구데이터가 매일 관리되는 방식을 개선하고 특히 공유 및 재사용을 촉진하면 연구의 풍부함과 신뢰성이 향상 될 수 있다”라고 기술하였으며, 여기에 그들은 다음과 같이 RDM이 중요하게 된 이유를 다음과 같이 제시하였다(Cox and Verbann 2018).

- 새로운 유형의 과학으로 인해 발생하는 ‘데이터의 홍수’
- 특정 분야의 연구 무결성에 대한 신뢰의 위기
- 오픈 데이터 운동

이상과 같이 RDM의 정의와 중요성을 살펴본 결과, “RDM은 연구 프로젝트의 연구 사이클 중 생성 및 수집되는 데이터의 조직, 저장, 공유, 접근, 보존 및 재사용의 방법을 다루는 것”으로 정의될 수 있다. 이러한 RDM의 효과는 다음과 같이 정리할 수 있다.

- 연구데이터의 손실 방지
- 다른 사람들과의 연구데이터 공유
- 검색, 접근 그리고 재사용 가능
- 연구에 투입되는 시간과 자원의 절약
- 연구의 오류 방지 및 품질 향상
- 연구 결과 검증 및 복제 가능
- 연구데이터 공유로 인한 새로운 발견 가능

## 2.2 연구데이터 관리 사례

본 절에서는 RDM 서비스 기관 4곳을 선택하여 사례를 분석하고자 한다. 4개의 기관 중 DCC는 일반적인 RDM 서비스 개발과 관련하여 정

책을 개발하기 위한 가이드라인으로 활용될 수 있고 ANDS, ICPSR 그리고 DataONE은 특정 학문 분야 및 다분야에 맞추어진 RDM 서비스 내용을 확인할 수 있어 선택하게 되었다.

2.2.1 DCC(Digital Curation Centre)  
Johns와 Pryor(2013)는 기관이 RDM 서비

스를 계획 및 구현할 수 있도록 RDM 서비스의 구성 요소와 프로세스를 설명하고 해당 서비스의 역할과 책임에 대해 기술하였다. 다음의 <표 2>는 DCC의 RDM 서비스 구성요소와 세부내용을 정리한 것이다.

DCC에서 제공하는 RDM의 첫 번째 구성요소는 'RDM 정책과 전략'이다. 해당 요소의 핵

<표 2> 영국 DCC의 RDM 서비스 구성요소와 내용

구성요소	내용
RDM 정책과 전략	<ul style="list-style-type: none"> <li>• 전략 개발</li> <li>• 정책 개발</li> <li>• 데이터 정책 및 전략 조정</li> </ul>
업무계획과 지속성	<ul style="list-style-type: none"> <li>• 전략에 따라 3년, 5년, 10년 계획을 다루는 단계별 사업 계획의 개발</li> <li>• 지출이 발생할 회계 연도의 표시와 함께 예상 비용 및 계획 지출 확인</li> <li>• 서비스 요금을 부과함으로써 비용을 회수할 수 있는지 고려</li> <li>• 투자에 대한 사례를 만드는 데 도움이 되는 비용/편익 분석 수행</li> <li>• 지속 가능성 이슈 및 관련 장기 비용 문제의 해결</li> </ul>
가이드, 교육 및 지원	<ul style="list-style-type: none"> <li>• 가이드와 헬프데스크</li> <li>• 이용자별 교육</li> <li>• 컨설팅 서비스</li> </ul>
DMP	<ul style="list-style-type: none"> <li>• DMP 서비스에 대한 필요성</li> <li>• DMP 템플릿 및 가이드</li> <li>• DMP 도구</li> <li>• 교육과 컨설팅</li> <li>• DMP 작성 담당</li> </ul>
현용 데이터 관리	<ul style="list-style-type: none"> <li>• 연구데이터 스토리지</li> <li>• 클라우드 스토리지 서비스</li> <li>• 학술 'dropbox' 서비스</li> <li>• RDM 플랫폼</li> </ul>
데이터 선정 및 수집	<ul style="list-style-type: none"> <li>• 보존할 데이터 선택 이유</li> <li>• 연구자가 데이터를 선택하고 기탁하도록 권장</li> <li>• 높은 수준의 가이드</li> <li>• 옹호 및 전공별 가이드</li> <li>• 기탁 도구</li> <li>• 기탁 동의</li> <li>• 선택 결정을 위한 기준 수립</li> </ul>
데이터 리퍼지토리	<ul style="list-style-type: none"> <li>• 기관 데이터 리퍼지토리</li> <li>• 외부 연구데이터 리퍼지토리</li> <li>• 연구자를 위한 관련 서비스 표시</li> </ul>
데이터 기술	<ul style="list-style-type: none"> <li>• 연구데이터셋을 기록하는데 필요한 메타데이터 정의</li> <li>• 연구데이터 보유 기록을 수집하고 표시하기 위한 시스템 설정</li> <li>• 가능한 경우 시스템을 통합하여 데이터 수집 및 메타데이터 생성을 기존 워크 플로우에 임베드</li> <li>• 국가 목록 또는 기타 관련 서비스에 포함시킬 메타데이터 공개</li> </ul>

심 활동에는 현재 위치와 전략에 대한 이해, 인프라와 서비스를 제공하기 위한 프로그램 작성, 핵심 RDM 원칙을 수립하기 위한 외부 동인 및 지역 상황에 따른 정책 작성, 정책 이행을 위한 옹호 활동과 시범 연구 마지막으로 합의를 얻고 지원을 확보하기 위한 광범위하고 반복적인 문의를 포함된다. 두 번째 요소는 '업무계획과 지속성'이다. 여기에서는 전략에 따른 중장기 계획 개발, 예상 비용 및 계획 지출 확인, 비용 회수 고려, 비용/편익 분석 수행 그리고 장기 비용 문제의 해결에 관한 주요한 핵심 활동이 포함된다. 세 번째 구성요소는 '가이드, 교육 및 지원'으로 다양한 이용자를 만족시키기 위해 다양한 수준의 지침과 훈련을 제공해야 한다. 여기에는 헬프데스크 또는 연락처를 통한 지원 제공, RDM 교육 그리고 심층 컨설팅 서비스가 포함된다. 네 번째 구성요소는 'DMP<sup>1)</sup>'이다. 여기에는 DMP 템플릿 또는 가이드라인 개발, 옵션 및 사례 제공, DMP 도구 그리고 교육 및 컨설팅 프로그램이 핵심 활동으로 제시된다. 다섯 번째는 '현용 데이터 관리'로서, 연구 활동 단계에서 데이터 관리를 지원하기 위한 서비스와 애플리케이션에 관한 내용을 기술하고 있다. 여기에는 추가 연구데이터 스토리지 제공을 위한 투자 사례 작성, 스토리지 할당 및 관리 절차 개발 그리고 데이터의 생성, 관리 및 공유를 지원하는 유연한 RDM 시스템 제공이 포함된다. 여섯 번째 '데이터 선정 및 수집'으로 하위 세부 내용으로 보존할 데이터의 선택 이유, 연구자

의 데이터 선택 및 기탁, 전공별 가이드 제시, 기탁 도구와 기탁 동의, 데이터 선택을 위한 기준 제시 등으로 구성되어 있다. 일곱 번째인 '데이터 리포지토리'는 기관 데이터 리포지토리, 외부 연구데이터 리포지토리 그리고 연구자를 위한 관련 서비스 표시 등 3개의 세부 내용으로 구성되어 있다. 이 요소의 핵심 활동에는 리포지토리 개발 및 유지, 기존 시스템과 리포지토리 조정, 외부 리포지토리 역할 결정 그리고 관련 서비스 안내 및 지원 등이 포함된다. 마지막 구성요소는 '데이터 기술'이다. 연구기관은 보유하고 있는 연구데이터에 대한 기록을 가지고 있으며, 이 메타데이터를 온라인으로 제공하여 데이터 검색 및 재사용을 지원해야 한다. '데이터 기술'의 핵심 활동에는 먼저 연구데이터셋을 기록하는데 필요한 메타데이터 정의, 연구데이터 보유 기록을 수집하고 표시하기 위한 시스템 설정, 시스템을 통합하여 데이터를 수집하고 메타 데이터 생성을 기존 워크플로우에 포함 그리고 국가 목록 또는 기타 관련 서비스에 포함시킬 메타데이터 공개 등이 제시되어 있다.

## 2.2.2 ANDS(Australia National Data Service)

ANDS는 2013년 1월에 「연구데이터 관리의 실제」라는 가이드를 통해 효과적인 연구데이터 관리를 지원할 수 있도록 연구데이터 관리의 모든 측면을 다루는 높은 수준의 제도적 프레임워크와 광범위한 사용 방법 안내서를 제공하였다.

1) DMP는 연구 프로젝트 중 및 프로젝트 완료 후 데이터를 처리하는 방법을 간략하게 설명하는 공식 문서를 말하며, DMP의 목표는 프로젝트가 시작되기 전에 데이터 관리, 메타 데이터 생성, 데이터 보존 및 분석의 여러 측면을 고려하는 것이다. DMP from Wikipedia [cited 2019. 7. 5.]  
 <[https://en.wikipedia.org/wiki/Data\\_management\\_plan](https://en.wikipedia.org/wiki/Data_management_plan)>

다음의 <표 3>은 ANDS가 제시한 연구데이터 관리 단계로서 연구 프로젝트 이전, 도중 및 이후에 연구기관이 연구데이터 관리에서 고려해야 할 주요 단계를 요약한 것이다(MERCURY PROJECT SOLUTIONS 2013).

<표 3> ANDS의 연구데이터 관리 단계별 내용

단계	내용
연구 전	연구데이터 관리 프레임워크
	위험 관리 계획
	연구데이터 관리 계획
	윤리적 허용
연구 중	교육 및 인도
	정책 준수 모니터링
	위험 모니터링 및 커뮤니케이션
	연구데이터 수집 및 분석
	메타데이터 생성
연구 종료 후	저장 및 접근
	연구데이터 출판
	연구데이터 등록
	지속적인 큐레이션
	사용 모니터링

ANDS에서 제공하는 연구데이터 관리 단계는 '연구 전', '연구 중' 그리고 '연구 종료 후' 3단계로 구성되며 각 단계별 세부 내용은 14개 항목으로 '연구데이터 관리 프레임 워크', '위험 관리 계획', '연구데이터 관리 계획', '윤리적 허용', '교육 및 유도', '정책 준수 모니터링', '위험 모니터링 및 커뮤니케이션', '연구데이터 수집 및 분석', '메타데이터 생성', '저장 및 접근', '연구데이터 출판', '연구데이터 등록', '지속적인 큐레이션' 그리고 '사용 모니터링' 등 14개 항목을 포함하고 있다.

### 2.2.3 ICPSR(Inter-university Consortium for Political and Social Research)

ICPSR은 이용자가 데이터를 재사용하고 연구 결과를 검증할 수 있도록 사회과학 데이터에 대한 저장, 큐레이트 및 접근을 제공한다. 이러한 서비스를 제공하기 위해 ICPSR은 데이터 관리 및 큐레이션 서비스를 여섯 개의 카테고리로 범주화하여 제시하였다. 여섯 개의 구성요소는 품질, 보존, 접근, 기밀유지, 인용, 도구 및 서비스로 되어 있으며 각각의 내용은 다음의 <표 4>와 같다(ICPSR 2019).

<표 4> ICPSR의 데이터 관리 및 큐레이션

구성요소	내용
품질	• 아카이빙을 위한 데이터 준비
보존	• 신뢰할 수 있는 디지털 리포지토리 • 디지털 보존 정책 및 계획
접근	• DMP • OSTP 데이터 접근 계획 가이드라인
기밀유지	• 기밀유지에 대한 ICPSR의 접근법
인용	• 데이터 인용 방법
도구 및 서비스	• 기밀유지 • 이용-제한 데이터 • 데이터 프로세싱 • DMP

ICPSR의 데이터 관리 및 큐레이션의 첫 번째 구성요소는 '품질'이다. '품질'의 세부내용은 아카이빙을 위한 데이터를 준비하는 내용으로 데이터 공유 및 아카이빙의 중요성 그리고 데이터 라이프 사이클이 제시되어 있다. 두 번째 구성요소인 '보존'은 신뢰할 수 있는 디지털 리포지토리와 디지털 보존 정책 및 계획으로 구성되어 있다. 연구데이터의 체계적인 관리와 활용을 위해서는 리포지토리가 필수적이다. 이러한 리포지토리는 연구자가 연구데이터를 기



탁, 공유, 접근 및 보존까지의 과정을 쉽게 이행되도록 구성되어 있다. 다음으로 ICPSR의 디지털 보존 정책 및 계획은 접근 정책 프레임워크, 컬렉션 개발 정책 그리고 기타 ICPSR 정책으로 구성되어 있다. 세 번째, '접근' 요소는 DMP와 'OSTP 데이터 접근 계획 가이드라인'으로 구성되어 있다. ICPSR은 현재 여러 연방 기관과 파트너 관계를 맺고 연방 펀딩 연구 결과에 공개적으로 접근할 수 있는 권한을 부여한다. 이들 기관은 ICPSR의 역량을 활용하여 데이터를 효율적으로 선별, 보존 및 배포한다. 네 번째는 '기밀유지'로서 ICPSR의 데이터는 기밀 및 개인 정보 보호 문제에 대해 선별되며, 민감한 데이터를 보호하고 배포하기 위해 엄격한 보호가 이루어진다. 연구에 참여하는 모든 연구자들은 연구 중에 생산되는 연구데이터의 보안에 대하여 많은 관심을 가지고 있다. 기술 진보에 따른 위험성 중에서 많은 사람들이 우려하는 부분이 바로 프라이버시 침해다. 특히, 사회과학 연구에서는 참여자의 개인 정보 보호가 연구 응답률을 높이는 방안이 될 수 있다. 다섯 번째 구성요소 '인용'으로는 ICPSR의 데이터 인용방법 등에 대한 내용이다. 적절한 인용은 연구데이터의 탐색, 재사용, 검증을 위한 복제, 인정을 위한 크레딧(credit) 그리고 사용 및 영향력을 측정하기 위한 추적 등을 가능하게 한다. 마지막 구성요소는 '도구 및 서비스'로서 기밀유지, 이용-제한 데이터, 데이터 프로세싱 그리고 DMP를 위한 것이다. 기밀 유지를 위한 도구 및 서비스는 공개 검토(Disclosure Review)와 텍스트 익명 헬퍼 도구이며, 이용 제한 데이터를 위한 서비스로는 물리적 데이터 영역과 가상 데이터 영역이 있다. 데이터 프로세싱에는

Nesstar Publisher, Open Science Framework 그리고 OpenRefine이 있으며 마지막으로 DMP를 작성하는 도구로서 DMPTool을 제공한다.

### 2.2.4 DataONE(Data Observation Network for Earth)

DataONE은 「Primer on Data Management」이라는 보고서를 통해 데이터 관리 계획을 개발하고 데이터를 효과적으로 작성, 구성, 관리, 설명, 보존 및 공유하는 방법에 대한 몇 가지 기본 데이터 관리 방법에 대해 설명하고 있다. 각 단계별 Data Life Cycle에 대한 세부 내용은 다음의 <표 5>와 같다(Strasser et al. n.d.).

<표 5> DataONE의 데이터 라이프 사이클 단계별 세부 내용

단계	내용
계획	• DMP 작성 안내
수집	<ul style="list-style-type: none"> <li>• 템플릿 작성</li> <li>• 데이터 파일 콘텐츠 기술</li> <li>• 일관된 데이터 구성 사용</li> <li>• 동일한 포맷 사용</li> <li>• 데이터명 등에 ASCII 문자사용</li> <li>• 안정적인 비-독점 소프트웨어 및 하드웨어 사용</li> <li>• 설명적인 파일 이름 할당</li> <li>• 원시 데이터를 원시로 유지</li> <li>• 매개 변수 테이블 작성</li> <li>• 사이트 테이블 작성</li> </ul>
보장	<ul style="list-style-type: none"> <li>• 품질 보증</li> <li>• 품질 관리</li> </ul>
기술	<ul style="list-style-type: none"> <li>• 디지털 컨텍스트 기술</li> <li>• 인력 및 이해 관계자</li> <li>• 과학 컨텍스트 기술</li> <li>• 파라미터 정보</li> </ul>
보존	<ul style="list-style-type: none"> <li>• 장기적 가치가 있는 데이터 식별</li> <li>• 적절한 정밀도를 사용하여 데이터 저장</li> <li>• 표준 용어 사용</li> <li>• 법률과 정책 고려</li> <li>• 기여와 출처</li> </ul>
탐색 통합	<ul style="list-style-type: none"> <li>• 데이터 검색, 통합, 분석 및 시각화를 지원하는 다양한 도구</li> </ul>
분석	<ul style="list-style-type: none"> <li>• 데이터를 통합, 분석 및 시각화하고 해당 프로세스에 사용된 정확한 단계를 문서화하는 복잡한 과학적 워크플로우 생성 및 관리</li> </ul>

첫 번째 단계인 '계획'에서는 DMP를 작성하기 위한 9개의 고려사항을 제시하고 있다. 고려사항에는 데이터 수집, 리포지토리 결정, 데이터 구성, 데이터 관리, 데이터 기술, 데이터 공유, 데이터 보존, 예산 고려 그리고 기관 리소스 탐색이 포함된다.

두 번째 단계인 '수집'에서는 향후 유용성을 보장할 수 있는 방식으로 수집 전에 방법과 문서를 신중하게 고려할 내용을 포함하고 있다. 여기에는 템플릿 작성, 데이터 파일 콘텐츠 기술, 일관된 데이터 구성 사용, 동일한 포맷 사용, 데이터명 등에 ASCII 문자 사용, 안정적인 비-독점 소프트웨어 및 하드웨어 사용, 설명적인 파일 이름 할당, 원시 데이터를 원시로 유지, 매개 변수 테이블 작성 그리고 사이트 테이블 작성 등이 제시되었다.

세 번째 단계인 '보장'은 데이터 수집, 입력 및 분석 중에 데이터에 대한 기본 품질 보증<sup>2)</sup> 및 품질 관리<sup>3)</sup>에 대한 수행을 의미한다.

네 번째 단계는 '기술'로서 포괄적인 데이터 도큐멘테이션(예: 메타데이터)을 나타내며 이것은 향후 데이터를 이해하는 데 중요하다. 즉, 데이터 파일의 컨텍스트, 데이터가 수집된 컨텍스트, 수행된 측정 및 데이터 품질에 대한 철저한 설명 없이는 데이터를 쉽게 탐색, 이해 또는 효과적으로 사용할 수 없을 것이다. 메타데이터는 가장 관련성이 높은 과학 커뮤니티에서 일반적으로 사용되는 메타데이터 포맷으로 생성해야 하며, 포괄적인 데이터 설명을 생성하면 다른 사람들이 해당 데이터를 탐색하고 이해하고 사용할 수 있다.

다섯 번째 단계는 '보존'이다. 데이터 보존을 위해서 해당 연구 분야에 익숙한 데이터 센터 또는 아카이빙 서비스를 이용해야 한다. 여기에는 공식 메타데이터를 준비하는 방법, 데이터를 보존하는 방법, 사용할 파일 포맷 및 향후 데이터 사용자에게 추가 서비스를 제공하는 방법에 대한 지침이 제공될 수 있다. 데이터 센터는 사용자 요구에 따라 데이터의 검색, 접근 및 배포를 지원하는 도구를 제공할 수 있다.

여섯 번째 단계 '탐색', 일곱 번째 단계 '통합' 그리고 여덟 번째 단계 '분석'은 데이터 검색, 통합, 분석 및 시각화를 지원하는 다양한 도구를 사용할 수 있다. 여기에서는 데이터를 통합, 분석 및 시각화하고 해당 프로세스에 사용된 정확한 단계를 문서화하는 복잡한 과학적 워크플로우의 생성 및 관리를 지원한다.

## 2.3 선행 연구

본 절에서는 RDM과 관련된 국내외 선행 연구를 분석하여 시사점을 도출하고자 한다.

먼저, 국내 연구데이터 관리서비스의 체계 및 운영을 개선하기 위하여 김지현(2014)은 미국의 연구 중심 대학도서관에서 제공하고 있는 연구데이터 관련 서비스를 9개의 구성요소를 기준으로 분석하였다. 기준으로 활용된 9개의 구성요소에는 'DMP 작성 지원', '데이터 파일 정리', '데이터 기술', '데이터 저장', '데이터 공유 및 접근', '데이터 보존', '데이터 인용', '데이터 관리 교육' 그리고 '데이터 지적재산권'이 포함된다. 또한, 미국 대학도서관의 연구데이터

2) 프로젝트의 목표를 충족시키는 방식으로 데이터가 생성되고 컴파일 되도록 하는 일련의 활동

3) 데이터의 문제를 식별하도록 설계된 테스트 또는 기타 활동

지원 사례를 연구하기 위하여 심원식(2016)은 버지니아, 존스 홉킨스, 퍼듀 그리고 일리노이 등 4개 대학을 대상으로 연구데이터 서비스 전달조직, 교육 서비스, 컨설팅 서비스 그리고 시스템 인프라를 상호 분석하였다. 해당 연구의 시사점으로 연구데이터에 대한 전문성 확보, 데이터 관리 교육 서비스 그리고 교내 IT 부서 등과의 협력을 통한 시스템 인프라 구축 등 세 가지 방안이 제시되었다. 뿐만 아니라 대학도서관 등 기관의 범주를 넘어 국가 차원의 연구데이터 관리 및 공유를 위한 로드맵을 제시하기 위하여 심원식(2015)은 영국의 RCUK(Research Councils UK) 산하 7개 연구비 지원기관과 산하 데이터센터, 그리고 관련 지원 기관의 기능을 중심으로 관계자와의 면담을 진행하였다. 면담을 통한 조사 분석 결과 해당 로드맵은 다음과 같이 제시되었다.

- 법제도/정책: 관련 법제 정비, 데이터관리 계획 정책 실시, 연구수행기관별 로드맵 구축, 기관별 데이터 정책 실시
- 인프라구축/운영: 연구데이터 인프라 구축
- 문화 정착: 이해관계자 공감대 형성, 국제적인 공조

연구데이터 관리서비스 구현 시 성공적인 서비스를 위한 고려사항을 도출하기 위하여 김성훈과 오삼균(2018)은 미국, 호주 그리고 독일의 대학도서관 및 기관의 담당자 8명을 대상으로 설문 및 면담을 진행하여 실제적인 고충과 전략적 성공방안을 조사 및 분석하였다. 분석 내용을 중심으로 국내 전문가의 의견을 수렴하여 연구데이터 관리서비스 구현 시 성공적인 서비스를 위한 고려 사항을 제시하였다. 고려사항에는

연구자에게 확실히 유익한 서비스 구성, 국가/대학/기관 차원의 전략 수립, 메타데이터 입력 주체 고려 및 필수 요소 최소화, 직원의 전문화, 이용자 요구분석을 통한 주요서비스 영역 선정 그리고 연구데이터와 연구결과물의 효과적인 연결방안 및 이용자와 유관기관과의 긴밀한 공조방안 마련 등이 포함되었다.

해외 선형 연구의 경우, 연구데이터 서비스를 제공하기 위한 도서관의 역할을 확인하기 위하여 Yu(2017)는 문헌과 ARL 그리고 ACRL로부터 수집된 서베이를 기반으로 데이터 관리 계획(DMP), 메타데이터 컨설팅과 도구 제공, 데이터 공유와 접근, 기관 리포지토리 제공 그리고 데이터 아카이빙 등 5가지의 연구데이터 서비스 분야를 도출하였다. 그는 연구자들의 데이터 보관 및 공유를 위해 로컬 리포지토리와 같은 오픈 액세스 리포지토리를 활용하도록 도서관이 안내하고 있으며, RDS 정책 및 인프라 개발에 대한 논의는 충분하지 않거나 거의 존재하지 않는다고 기술하였다. 또한 연구데이터 관리 서비스를 개발하기 위하여 Perrier와 Barnes(2018)는 포커스 그룹으로부터 원활한 인프라, 데이터 보안, 기술 및 지식 개발 그리고 데이터 공개에 대한 불안 등 4가지의 테마를 확인하였다. 특히, 그들은 도서관이 연구데이터 관리 서비스 개발 우선 순위를 다음과 같이 데이터 관리와 연구 지원 서비스 2가지 분야로 구분하여 나타냈다(〈표 6〉 참조).

연구자가 생산하는 데이터를 지원하기 위한 도구(Support Your Data)를 개발하기 위하여 Borghi 등(2018) 연구자들이 연구데이터 관리를 자체 평가할 수 있는 항목과 RDM 관련 서비스 및 정보를 제시하였다. 또한 기관 상황에

맞게 수정할 수 있는 보완 가이드 세트 6개(계획, 조직화, 저장, 준비, 분석, 공유)로 구성된 연구데이터 관리 가이드를 개발하였다.

〈표 6〉 연구데이터 관리 서비스 개발 우선 순위

분야	서비스 우선순위
데이터 관리	(1) 안전한 데이터 저장 (2) 데이터 관리 계획(DMP) (3) 파일 정리 (4) 데이터 기술 (5) 장기간 접근을 위한 데이터 보존 (6) 데이터 공유를 위한 데이터 리퍼지토리 확인
연구 지원 서비스	(1) 자금 지원 기관 요구 사항 (2) 저널 요구 사항 (3) 지적재산권 (4) 데이터셋 인용 (5) 리포지토리에서 데이터 찾는 방법 (6) 동의서 (7) 기밀 유지

이상과 같이 7건의 국내외 선행 연구를 살펴본 결과 국내 연구는 주로 연구데이터 관리 현황과 서비스의 구성요소 그리고 서비스 실행을 위한 로드맵에 관한 것이었고 국외 연구는 연구데이터 관리를 위한 대학도서관의 역할과 서비스 개발 그리고 연구데이터를 관리하기 위한 가이드 제시 등으로 나타났다. 이러한 선행 연구는 연구데이터 서비스를 위한 구성 요소(DMP, 파일 정리 등)만을 제시하여 RDM 시스템 설계 시 구체적인 가이드로 사용될 수 없는 한계를 내포하고 있다. 이에 본 연구에서는 기존 연구에서 시도되지 않은 연구데이터 관리를 위한 서비스의 하위 요소를 도출하여 RDM 시스템 설계 시 가이드라인으로 적용할 수 있다는 게 차이점이라고 할 수 있다.

### 3. 연구데이터 관리의 기능별 세부요소 조사 및 분석

본 장에서는 Geoscience 분야의 연구데이터 특성 및 관련 기관의 연구데이터 관리 사례에 대하여 분석하고자 한다.

#### 3.1 Geoscience 분야 연구데이터 특성

본 절에서는 문헌 조사를 통해 Geoscience 분야의 연구데이터 특성에 대하여 살펴보고자 한다.

Huadong(2017)의 연구에 의하면, Geoscience 분야 데이터는 다른 과학데이터와 다르게 대규모, 다중 소스, 다중 시간, 다중 규모, 고차원적, 높은 복잡도, 비정형적, 비구조적인 특징을 갖는다고 하였다. 해당 특징은 다음과 같다.

- 대규모: 해상도가 높고, 동적이며, 여러 대역으로 구성. 높은 데이터 획득 비율과 짧은 사이클을 가짐
- 다중 소스: 이미지 매커니즘과 모델이 매우 다양하기 때문에 데이터 소스와 획득 방법이 다양함
- 다중 시간: 샘플링 간격이 짧고 데이터 수집 빈도가 높음
- 고가치: 생태 환경, 토지 자원, 자연 재해 및 기타 Geoscience 분야 연구의 중요 가치를 가짐
- 고차원 및 고도로 복잡: 지역 규모가 국가 및 세계 규모이며 시간 단위가 초에서 수천 년에 이름
- 기타: 각 환경(각 시스템마다 다른 시공간 스케일)에서 획득한 Geoscience 분야 데이터는 서로 다른 규칙과 특성이 존재

한중규(2019)는 지질자원 연구데이터의 특징에 대하여 데이터 획득 방법, 데이터의 종류 및 형태, 데이터 파일 포맷 등이 다양하게 나타난다고 주장하였다. 다음의 <표 7>은 해당 연구에서 기술된 데이터의 특징을 정리한 것이다.

이러한 지질자원 연구데이터를 전공별로 구분하면 지질, 자원, 해양, 물리, 화학, 생물, 금속 등 다양하게 나타나며 육지/바다, 지질연대, 시기 등 시공간 데이터가 많은 것이 특징이라고 할 수 있다. 데이터 획득도 연구자가 직접 야외에서 관찰 또는 채취한 지질시료를 분석하거나 선박, 차량, 항공기 등 자료획득 플랫폼을 이용하는 등 다양한 방법으로 이루어지고 있다. 또한 데이터의 종류 및 형태와 파일 포맷의 경우 물리적/디지털 등 정형/비정형 데이터가 혼재되어 있다. 특히, 지질분야 연구데이터는 데이터 취득을 위한 인력, 시간 그리고 예산이 많이 필요하며 미래예측에 필요한 과거 데이터의 보존 및 활용이 중요하다고 언급하고 있다.

### 3.2 Geoscience 분야 연구데이터 관리 사례 분석

본 절에서는 Geoscience 분야의 연구데이터 관리를 위한 기능별 세부요소를 도출하기 위하여 관련 기관의 연구데이터 관리 사례를 분석하였다. 분석 대상 기관은 다음과 같다.

- BGS(British Geological Survey)
- USGS(US Geological Survey)
- CSIRO(Commonwealth Scientific and Industrial Research Organisation)
- IEDA(Interdisciplinary Earth Data Alliance)

#### 3.2.1 BGS(British Geological Survey)

BGS는 영국 정부를 위한 공공 과학, 지구 및 환경 프로세스를 이해하기 위한 공공 펀딩 기관이다. NGDC(National Geoscience Data Centre)는 이러한 BGS에서 운영하는 영국 최고의 Geoscience 분야 국가 컬렉션으로서 Geoscience 분야 데이터 및 정보를 수집·보존한다(Garry R. Baker n.d.).

<표 7> 지질자원 연구데이터의 특징

특징	세부 내용
전공	지질, 자원, 해양, 물리, 화학, 생물, 금속 등
시공간 데이터	육지/바다, 지질연대, 시기
데이터 획득 방법	<ul style="list-style-type: none"> <li>• 사람이 직접 야외에서 암석/지질구조/지형 등을 관찰하고 야장에 기록</li> <li>• 야외에서 암석, 토양, 지하수, 시추코어 등 지질시료를 채취하여 실험실에 가져와 분석장비를 이용하여 분석</li> <li>• 야외에서 탐사/측정 장비를 이용하여 자료를 획득</li> <li>• 다양한 자료획득 플랫폼: 선박, 차량, 항공기, 드론, 위성, 고정설치장비, 휴대장비 등</li> </ul>
데이터의 종류 및 형태	<ul style="list-style-type: none"> <li>• 물리적: 육상/해저 시추코어, 암석, 토양, 지하수 등 시료</li> <li>• 디지털: 스프레드시트, 사진, 영상, 이미지, 지도(GIS), 모델, 소프트웨어 등</li> </ul>
데이터 파일 포맷	정형/비정형 데이터 혼재

다음의 <표 8>은 BGS가 제공하는 RDM의 세부 내용을 정리한 것으로 크게 데이터 기탁과 데이터 접근 두 가지로 나눌 수 있다(BGS 2019).

먼저, 데이터 기탁은 NGDC에 데이터를 기탁하기 위한 가이드라인에 따라 좋은 데이터를 어떻게 기탁하는지에 대한 체크리스트와 허용되는 파일 포맷을 제공하여 실제 연구자가 데이터를 제출할 수 있도록 해당 기능을 제공하고 있다. 최근 과학 문헌에 게재된 논문과 그 기초가 되는 데이터 사이의 강한 연계, 그

리고 인용으로 데이터 수집을 보상하는 메커니즘에 대한 과학계의 요구가 증가하고 있다. 인용 데이터는 연구결과에 대한 크레딧, 데이터 저널 출판, 데이터 가치 공개 그리고 데이터 이용 보장 등에 대한 근거로서 활용될 수 있다. 데이터가 인용되면 해당 연구결과에 대한 신뢰도가 높아지고, 데이터를 가지고 데이터 저널에 출판할 수 있으며 해당 데이터의 가치가 공개되며 그리고 데이터의 이용을 보장할 수 있다.

<표 8> BGS의 RDM 세부 내용

구분		세부 내용
데이터 기탁	NGDC에 데이터 기탁	<ul style="list-style-type: none"> <li>• NGDC 데이터 기탁 가이드라인</li> <li>• NGDC 검토과제</li> <li>• 좋은 데이터 기탁 가이드라인</li> <li>• 허용할 수 있는 포맷 리스트</li> <li>• NGDC 데이터 가치 체크리스트</li> </ul>
	NGDC에 시추(drill) 의향 전달	<ul style="list-style-type: none"> <li>• 시추공을 뚫거나 우물을 팠 경우 시추 의향서를 작성해야 함</li> </ul>
	데이터 인용	<ul style="list-style-type: none"> <li>• 연구결과에 대한 크레딧</li> <li>• 데이터 저널 출판</li> <li>• 데이터 가치 공개</li> <li>• 데이터 이용 보장</li> </ul>
	데이터 관리	<ul style="list-style-type: none"> <li>• 지질 데이터 관리</li> <li>• 시추공 데이터</li> <li>• 경제 광물 및 건축 석재</li> <li>• 지형/지질 사진의 국립 기록 보관소</li> <li>• 지반 데이터</li> <li>• 암석 분류 체계 및 기타 Geoscience 분야 어휘</li> <li>• 층서 데이터   암석 단위 사진</li> </ul>
	펀딩을 받은 NERC 연구자를 위한 정보	<ul style="list-style-type: none"> <li>• NERC의 재정 지원을 받는 연구원들이 NGDC를 통한 데이터 기탁에 관한 정보 제공</li> </ul>
	NERC 데이터 정책	<ul style="list-style-type: none"> <li>• NERC의 재정 지원을 받는 모든 사람들이 수집한 데이터를 관리하기 위한 데이터 정책</li> </ul>
데이터 접근	기탁 데이터 탐색	<ul style="list-style-type: none"> <li>• NGDC 및 NGR(National Geological Repository)에 의해 수집된 데이터 탐색</li> </ul>
	데이터 컬렉션	<ul style="list-style-type: none"> <li>• BGS의 가장 많이 이용하는 데이터베이스 및 용어에 대한 탐색, 열람 그리고 다운로드</li> </ul>
	오픈 지오사이언스	<ul style="list-style-type: none"> <li>• BGS 데이터 웹사이트의 일부 NGDC 데이터 컬렉션 접근</li> </ul>
	OGL (Open Government Licence)	<ul style="list-style-type: none"> <li>• OGL은 넓은 범위의 공공 부문 정보를 무료로 재사용할 수 있는 간단한 이용 약관 세트</li> </ul>

NGDC의 두 번째 기능인 '데이터 접근'은 기탁된 데이터에 대한 탐색, 데이터 컬렉션, 오픈 Geoscience 그리고 데이터 라이선스를 위한 OGL로 구성되어 있다. 먼저 기탁된 데이터에 대한 탐색은 NGDC 및 NGR에 의해 수집된 데이터 탐색을 제공한다. 데이터 컬렉션은 BGS가 보유한 아카이브와 데이터 컬렉션에 대한 접근을 제공하고 있으며 이 정보의 대부분은 색인 및 데이터베이스화 되어 있어 온라인으로 데이터 수집을 검색하고 다운로드 할 수 있다. 여기에는 시추공 재료, 암석 및 화석 수집, BGS 암석 분류 체계 및 사전과 관련된 데이터가 포함된다. 다음으로 오픈 지오사이언스는 지도 확인, 데이터 다운로드, 스캔, 사진 및 기타 정보를 볼 수 있는 서비스이다. 마지막으로 OGL은 BGS 라이선스 정책을 의미하며, 공공 부문 정보를 무료로 재사용할 수 있는 간단한 이용 약관 세트를 가리킨다. 여기에는 몇 가지 조건만으로 이 라이선스에 따라 이용 가능한 정보를 자유롭게 이용할 수 있도록 그리고 유연하게 재사용할 수 있도록 방향성을 제시하고 있다.

### 3.2.2 USGS(U.S Geological Survey)

USGS는 체계적으로 데이터를 관리하기 위한 과학 데이터 라이프 사이클 모델을 개발했다. 이 모델은 데이터 관리 활동이 프로젝트 워크플로우와 관련되는 방식을 설명하고 기대치를 이해하는 데 도움을 준다. 또한, 해당 모델을 연구 활동에 적용함으로써 데이터 제품이 잘 설명되고 보존되고 접근 가능하며 재사용에 적합하도록 보장 할 수 있다. 여기에 USGS가 과학 데이터 관리를 위한 정책과 관행을 평가 및 개선하고 새로운 도구와 표준이 필요한 영역을 식

별할 수 있는 구조로도 사용한다(Faundeen, J.L. et al, 2013).

다음의 <표 9>는 USGS의 데이터 라이프 사이클을 기술한 것이다.

USGS의 데이터 라이프 사이클은 크게 순차적 단계와 모든 라이프 사이클 활동으로 구분할 수 있다. 먼저, 순차적 단계는 계획, 수집, 프로세스, 분석, 보존 그리고 출판/공유로 6단계로 구성되며 모든 라이프 사이클 활동에는 기술(메타데이터, 도큐멘테이션), 품질 관리 그리고 백업 및 보안 등 3개의 활동을 포함한다.

순차적 단계의 첫 번째인 '계획'은 데이터 관리 계획을 작성하고 중요한 계획 활동에 대해 학습하는 단계로서 프로젝트를 시작하기 전에 라이프 사이클 동안 데이터를 관리하는 방법을 계획하는 것이 중요하다. 여기에는 DMP 작성, 데이터 관리자의 역할 및 책임, 데이터 공유 계약 개발 그리고 데이터 접근 제어 및 저작권 이해 등이 포함된다. 두 번째 단계인 '수집'은 새로운 데이터의 수집 및 생성 또는 기존 데이터를 확보하는 단계로서 데이터 수집 방법, 데이터 수집을 위한 보안 요구 사항, 데이터 표준 준수, 데이터 템플릿 사용, 파일 유형 선택 그리고 파일 및 데이터 구성 등이 세부 내용을 제시된다. 세 번째 단계인 '프로세스'는 신규 또는 이전에 수집한 데이터의 입력 준비와 관련된 데이터 검증, 데이터 요약, 데이터 변환, 데이터 통합, 데이터 하위 설정 그리고 데이터 파생 등이 포함된다. 네 번째 단계인 '분석'에는 처리된 데이터의 탐색 및 해석과 관련된 통계 분석, 시각화, 공간 분석, 이미지 분석, 모델링 그리고 해석 등이 제시된다. 다섯 번째 단계인 '보존'에는 데이터의 장기적인 생존 및 접근성을 보장

〈표 9〉 USGS의 데이터 라이프 사이클

단계	세부 내용	
계획	<ul style="list-style-type: none"> <li>• DMP 작성</li> <li>• 데이터 관리자의 역할 및 책임</li> <li>• 데이터 공유 계약 개발</li> <li>• 데이터 접근 제어 및 저작권 이해</li> </ul>	
수집	<ul style="list-style-type: none"> <li>• 데이터 수집 방법</li> <li>• 데이터 수집을 위한 보안 요구 사항</li> <li>• 데이터 표준 준수</li> <li>• 데이터 템플릿 사용</li> <li>• 파일 유형 선택</li> <li>• 파일 및 데이터 구성</li> </ul>	
프로세스	<ul style="list-style-type: none"> <li>• 데이터 검증</li> <li>• 데이터 요약</li> <li>• 데이터 변환</li> <li>• 데이터 통합</li> <li>• 데이터 하위 설정</li> <li>• 데이터 파생</li> </ul>	
분석	<ul style="list-style-type: none"> <li>• 통계 분석</li> <li>• 시각화</li> <li>• 공간 분석</li> <li>• 이미지 분석</li> <li>• 모델링</li> <li>• 해석</li> </ul>	
보존	<ul style="list-style-type: none"> <li>• 데이터 아카이빙</li> <li>• 데이터 이관 및 폐기</li> <li>• 데이터 리포지토리</li> </ul>	
출판/공유	<ul style="list-style-type: none"> <li>• 데이터 공유 이유</li> <li>• USGS 데이터 릴리스 프로세스</li> <li>• 민감한 데이터</li> <li>• DOI</li> <li>• 데이터 인용</li> <li>• 데이터 목록</li> </ul>	
Cross-Cutting 요소	기술 (메타데이터, 도큐멘테이션)	<ul style="list-style-type: none"> <li>• 메타데이터 생성</li> <li>• 데이터 사전의 중요성</li> <li>• 데이터 검색에 키워드 사용</li> </ul>
	품질 관리	<ul style="list-style-type: none"> <li>• 품질 보증 계획</li> <li>• 품질 보증</li> <li>• 품질 관리</li> <li>• 데이터 품질 문서화</li> </ul>
	백업 및 보안	<ul style="list-style-type: none"> <li>• 데이터 백업 및 보안은 우발적인 데이터 손실, 손상 및 무단 접근으로 부터 데이터를 보호</li> </ul>



하기 위해 사용되는 조치 및 절차로서 데이터 아카이빙, 데이터 이관 및 폐기 그리고 데이터 리포지토리 등으로 구성되어 있다. 여섯 번째 단계인 ‘출판/공유’는 데이터를 출판하고 공유하는 것으로 동료들이 검토한 전통적인 저널 기사를 게재하는 것과 마찬가지로 연구 과정에서 중요하고 필요한 단계이다. 여기에는 데이터 공유 이유, USGS 데이터 공개 프로세스, 민감한 데이터, DOI, 데이터 인용 그리고 데이터 목록이 포함된다.

다음은 데이터 라이프 사이클 동안 순차적으로 발생되지는 않지만 데이터 라이프 사이클 모든 단계에서 지속적으로 수행되어야 할 활동인 ‘기술(메타데이터, 도큐멘테이션)’, ‘품질 관리’ 그리고 ‘백업 및 보안’ 등에 대한 내용이다. 먼저, 데이터 기술은 메타데이터와 도큐멘테이션을 말하며, 데이터 라이프 사이클 동안 데이터에 대한 조치를 반영하도록 도큐멘테이션의 작성과 데이터 업데이트를 명시하고 있다. 다음으로 품질 관리는 과학 데이터 라이프 사이클의 모든 단계에서 데이터를 올바르게 수집, 처리, 처리, 사용 및 유지 관리하기 위해 프로토콜 및 방법을 사용하는 프로세스로서 여기에는 품질 보증 계획, 품질 보증, 품질 관리 그리고 데이터 품질 문서화 등이 제시된다. 마지막으로 데이터 백업 및 보안은 우발적인 데이터 손실, 손상 및 무단 접근으로부터 데이터를 보호하는 기능을 수행한다.

### 3.2.3 CSIRO(Commonwealth Scientific and Industrial Research Organisation)

CSIRO는 데이터의 저장 및 관리, 데이터 출판, 데이터 공유, DMP, 데이터 탐색, 원칙 및

정책 그리고 교육 등의 서비스를 제공하여 연구 데이터 관리를 지원하고 있다. 다음의 <표 10>은 CSIRO의 연구데이터를 관리하기 위한 서비스를 기술한 것이다(CSIRO 2019).

<표 10> CSIRO의 연구데이터 관리 서비스 개요

서비스	세부 내용
데이터 저장 및 관리	<ul style="list-style-type: none"> <li>•DMP의 주요 이슈</li> <li>•네이밍과 데이터 기술</li> <li>•포맷</li> <li>•파일 네이밍과 버전 관리</li> <li>•학문별 가이드</li> <li>•연구데이터 스토리지</li> </ul>
데이터 출판	<ul style="list-style-type: none"> <li>•데이터 출판</li> <li>•이익</li> <li>•데이터 기술</li> <li>•데이터 인용</li> <li>•민감 데이터/데이터 윤리</li> <li>•DOI</li> </ul>
데이터 공유	<ul style="list-style-type: none"> <li>•데이터 공유</li> <li>•데이터 접근 포털</li> <li>•저널 데이터 공유 정책</li> <li>•펀더 데이터 출판 및 공유 정책</li> </ul>
연구데이터 관리 계획	<ul style="list-style-type: none"> <li>•연구데이터 관리자</li> </ul>
데이터 탐색	<ul style="list-style-type: none"> <li>•데이터 탐색</li> <li>•디렉토리 및 데이터 리포지토리</li> <li>•데이터 탐색 방법 - 11가지 팁</li> </ul>
데이터 접근 포털	<ul style="list-style-type: none"> <li>•데이터 접근 포털</li> </ul>
원칙 및 정책	<ul style="list-style-type: none"> <li>•호주 정부 공공 데이터 정책 선언</li> <li>•호주 오픈 데이터 현장</li> <li>•NHMRC 오픈 액세스 정책</li> <li>•Data.gov.au 정책</li> </ul>
교육	<ul style="list-style-type: none"> <li>•연구자를 위한 정보자원</li> <li>•도메인 교육(ANDS)</li> <li>•데이터 접근 포털 이용자 가이드</li> <li>•데이터 사이언스 - 온라인 코스</li> </ul>

첫째, ‘데이터 저장 및 관리’는 DMP의 주요 이슈, 네이밍과 데이터 기술, 포맷, 파일 네이밍

과 버전 관리, 학문별 가이드 그리고 연구데이터 스토리지로 구성된다. 두 번째 '데이터 출판'에는 데이터 출판의 이점, 데이터 기술, 데이터 인용, 민감 데이터/데이터 윤리 그리고 영구 식별자인 DOI에 대한 세부내용을 포함하고 있다. 세 번째인 '데이터 공유'에서는 데이터 공유의 필요성과 장점 그리고 리포지토리를 통한 데이터 저장에 대해 기술하고 있다. 네 번째 'DMP'는 연구데이터를 관리하기 위한 통합 계획 도구인 RDP(Research Data Planner)에 대하여 기술하고 있다. 이러한 RDP는 위험을 감소하게 하고 팀워크의 시간 절약, 연구 결과 개선 그리고 현재와 미래의 데이터를 쉽게 찾고, 액세스하고, 저장하고, 작업할 수 있도록 지원한다.

다섯 번째인 '데이터 탐색'은 실제 데이터를 탐색하도록 하는 포털과 리포지토리 등에 대한 목록을 제시한다. 여섯 번째인 데이터 접근 포털(DAP)은 CSIRO의 기본 데이터 리포지토리이며, re3data는 연구데이터 리포지토리 등록 사이트, OpenDOAR는 오픈 액세스 리포지토리 디렉토리 그리고 RDA(Research Data Australia)는 호주의 연구데이터 포털을 가리킨다. 일곱 번째는 '데이터 관리 원칙과 정책'에 대한 내용으로 여기에는 '호주 정부 공공 데이터 정책 선언', '호주 오픈 데이터 헌장', 'NHMRC 오픈 액세스 정책', 'Data.gov.au 정책', '책임에 대한 호주 법령' 그리고 '호주 주정부 공개 데이터 정책'을 포함하고 있다. 마지막으로 '교육'은 연구데이터 관리에 관한 교육자료로서 여기에는 연구자를 위한 정보자원, 도메인 교육, 데이터 접근 포털 이용자 가이드, 데이터 사이언스(온라인 코스) 그리고 도서 등으로 구성되어 있다. 이 중 '연구자를 위한 정보자원'은 연구자, 저자 및 저

널이 데이터 라이프 사이클 동안 데이터를 관리하는 데 도움이 되는 리소스, 도구 및 가이드 라인을 제공한다.

### 3.2.4 IEDA(Interdisciplinary Earth Data Alliance)

IEDA는 미국 국립과학재단의 지원을 받아 세계 지구 화학 및 해양 지구 과학 연구를 위한 주요 커뮤니티 데이터를 수집하여 광범위한 관측 분야와 분석 데이터 유형의 보존, 탐색, 검색 및 분석을 지원하고 있다. 다음의 <표 11>은 IEDA의 연구 프로세스 진행에 따라 지원되는 도구를 중심으로 기술된 데이터 관리 개요를 요약한 것이다.

IEDA의 연구데이터 관리는 프로세스에 따라 구분되며 크게 '연구 전', '연구 및 분석' 그리고 '합성 및 출판'으로 구분된다.

첫째, '연구 전' 단계에서는 IEDA 데이터 탐색 도구가 제공되는데 이는 과학적 질문과 관련된 사용 가능한 데이터를 검색할 수 있도록 한다. 다음으로 IEDA DMP 도구는 데이터를 문서화하고 보존하는 전략을 제시하며, 이것은 데이터 관리 프로세스의 초기 단계에 사용할 것을 강력하게 권장되며, DMP는 궁극적으로 데이터 준수를 입증하는 데 중요한 구성 요소로 제시되고 있다. 두 번째, '연구 및 분석' 단계에서는 실제 데이터를 등록하거나 분석 도구를 통해 해당 데이터를 분석하는 내용이 포함된다. IEDA는 데이터를 효율적이고 일관성 있게 문서화하는 방법에 대한 템플릿, 모범 사례 및 제안을 제공하며, 연구 완료시 NSF와 IEDA DMP를 연결하여 데이터 준수 보고를 용이하게 해준다. 또한, IEDA에 데이터셋과 샘플

〈표 11〉 IEDA의 데이터 관리 개요

프로세스	항목(도구 등)	
연구 전	IEDA 데이터 탐색 도구	
	IEDA DMP 도구	
연구 및 분석	IEDA 연락	
	IEDA에 데이터셋 및 샘플 등록	<ul style="list-style-type: none"> <li>▪ 샘플 기반 데이터                             <ul style="list-style-type: none"> <li>- 샘플 메타데이터 등록 및 unique sample identifier(IGSN)</li> <li>- 분석 데이터 용 템플릿 다운로드</li> <li>- EarthChem Library에 데이터 제공</li> </ul> </li> </ul>
		<ul style="list-style-type: none"> <li>▪ 센서 기반 데이터                             <ul style="list-style-type: none"> <li>- 센서 데이터 파일 및 지원 메타데이터를 MGDS에 제공</li> </ul> </li> </ul>
	IEDA 분석 도구	<ul style="list-style-type: none"> <li>▪ GeoMapApp</li> </ul>
합성 및 출판	IEDA에 최종 데이터셋 등록	
	데이터 출판	
	IEDA 데이터 준수 보고 도구	

플을 등록할 수 있도록 가이드를 제공하는데 샘플-기반 데이터의 경우, 샘플 메타데이터를 등록하고 고유한 샘플 식별자(International Geo Sample Number, IGSN)를 부여할 수 있는 사이트(<http://www.geosamples.org>)와 분석 데이터용 템플릿을 다운로드할 수 있는 사이트(<http://www.earthchem.org/data/templates>) 그리고 EarthChem Library에 분석 데이터를 제공하는 방법을 제시하고 있다. 다음으로 센서 기반 데이터의 경우, 센서 데이터 파일(예: 지구 물리학적 데이터) 및 지원 메타데이터를 MGDS에 제공하도록 안내하고 있다. 여기에 'GeoMapApp'이라는 IEDA 분석 도구를 제시하여 사용 가능한 다른 데이터셋의 범위 내에서 지리 공간 데이터를 분석하도록 지원한다. 세 번째, '합성 및 출판' 단계는 데이터셋의 등록 및 출판 그리고 데이터 준수를 보고하기 위한 도구를 제공한다. IEDA는 DOI® 시스템에 과학 데이터를 등록하는 데이터 출판 서비스를 제공하여 데이터셋을 연구자에게 저자로 귀속

하여 출판할 수 있도록 한다. 다음으로 IEDA 데이터 준수 보고 도구는 NSF 관리 번호(NSF Award Number)를 기반으로 데이터 준수 보고서를 신속하게 생성하여 데이터가 IEDA 시스템에 등록되어 있고 NSF 데이터 정책을 준수하고 있음을 보여주는 역할을 한다.

#### 4. Geoscience 연구데이터 관리의 기능별 세부요소 구성과 중요도

본 장에서는 3장에서 분석한 8개 기관의 RDM 사례를 중심으로 RDM 기능별 세부요소를 제시하고 그 세부요소의 중요도를 평가하여 우선순위를 도출하고자 한다.

##### 4.1 기능별 세부요소 구성

본 절에서는 RDM 기능별 세부요소를 제시하기 위하여 3장에서 분석한 8개 기관의 RDM

관련 내용을 기능별 세부요소로 제시하였다. RDM 기능은 김지현(2014)의 『대학도서관의 연구데이터관리서비스에 관한 연구』에서 제시한 RDM 서비스의 9개 구성 요소를 활용하였다. 김지현(2014)의 9개 구성 요소를 활용한 이유는 4편의 선행논문에서 제시된 RDM 구성 요소를 비교하여 2편 이상의 연구에서 제시한 구성 요소를 모두 포함하여 기준의 타당성을 포함하였다고 판단하였기 때문이다.

#### 4.1.1 DMP 작성 지원

8개 기관의 분석 내용 중 'DMP 작성 지원'과 관련한 세부요소를 도출한 결과 18개 요소

가 제시되었다. 18개 요소 중 동일하다고 판단되는 요소끼리는 합하고 세부요소로 적절치 않다고 생각되는 것은 제거하였다. 이러한 방식으로 제시된 'DMP 작성 지원'은 다음의 <표 12>와 같이 6개의 세부요소로 정리되었다.

#### 4.1.2 데이터 파일 정리

'데이터 파일 정리'와 관련하여 8개 기관의 내용을 분석한 결과 22개의 세부요소가 도출되었다. 유사하거나 동일하다고 판단되는 요소는 합하고 적절치 않는 요소는 제거하는 등 이러한 과정을 거쳐 '데이터 파일 정리'는 다음의 <표 13>과 같이 7개의 세부요소로 구성하였다.

<표 12> DMP 작성지원의 세부요소

RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
DMP 작성지원								
DMP 작성을 위한 템플릿 및 가이드 제시	○		○		○	○		
DMP 작성 도구 제시	○					○		○
DMP 작성을 위한 교육과 컨설팅 제공	○					○		
DMP 구성요소 안내		○	○	○				
DMP 작성을 위한 샘플 및 체크리스트 제시		○		○	○	○	○	
DMP 평가			○					

<표 13> 데이터 파일 정리의 세부요소

RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
데이터 파일 정리								
권장하는 파일 포맷 제공		○	○	○	○	○	○	○
파일 네이밍 규칙 및 모범사례 제시		○	○	○	○	○	○	
데이터 버전 관리 방법 제공		○	○	○	○		○	
데이터 및 파일 구조 제시			○					
데이터 표준 제시						○		
데이터 모델 정의				○				
데이터 수집, 공개, 보존 시 파일 포맷에 관한 모범사례 제공						○		

4.1.3 데이터 기술

‘데이터 기술’과 관련하여 8개 기관의 내용을 분석한 결과 24개의 세부요소가 도출되었다. 24개의 세부요소 중에서 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 기술’은 다음의 <표 14>와 같이 10개의 세부요소로 구성하였다.

4.1.4 데이터 저장

‘데이터 저장’과 관련하여 8개 기관의 내용을 분석한 결과 16개의 세부요소가 도출되었다. 16개의 세부요소 중 유사하거나 동일한 요소를 합한 결과 ‘데이터 저장’은 세부요소는 다음의 <표 15>와 같이 12개 세부요소로 구성하였다.

<표 14> 데이터 기술의 세부요소

데이터 기술 \ RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
메타데이터 표준 및 스키마 제공	○	○	○	○	○	○	○	
메타데이터를 위한 시스템 설정	○							
메타데이터 생성		○		○	○	○		○
메타데이터 공개	○				○			
메타데이터 작성 모범사례 제시			○			○	○	○
메타데이터 수집, 업데이트 및 유지 관리		○		○				
데이터와 메타데이터의 영구 연결을 지원하는 영구 식별자 제시		○						
메타데이터 레코드 작성 도구 안내				○		○		
메타데이터 유효성 검증 도구 안내						○		
메타데이터 검토 체크리스트 제시						○		

<표 15> 데이터 저장의 세부요소

데이터 저장 \ RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
연구 데이터 스토리지 할당 및 관리 절차 개발	○							
프로젝트 초기에 데이터 저장 계획 수립		○						
연구 데이터 저장을 위한 스토리지 제공	○							
저장할 연구 데이터 사본의 개수와 동기화 방법 제시			○					
사본 저장 방법과 위치 지정			○					
클라우드 기반 서비스에 저장된 데이터 전송 시스템의 백업을 위한 사본 사이트 보유			○					
정전시 백업 사이트에서 데이터 다운로드 서비스 제공			○					
스토리지 솔루션 비교 기준 제시		○					○	
데이터 백업시 무결성 및 접근성 보장				○	○			
저장 매체의 안정성 보장				○				
데이터 백업 모범 사례 제시 및 도구 안내				○	○	○		
데이터 보안 모범 사례 제시				○		○		

4.1.5 데이터 공유 및 접근

‘데이터 공유 및 접근’과 관련하여 8개 기관의 내용을 분석한 결과 25개의 세부요소가 도출되었다. 도출된 25개의 세부요소 중에서 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 공유 및 접근’은 다음의 <표 16>과 같이 12개의 세부요소로 구성하였다.

4.1.6 데이터 보존

‘데이터 보존’과 관련하여 8개 기관의 내용을 분석한 결과 26개의 세부요소가 도출되었다. 도출된 26개 중 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 보존’은 다음의 <표 17>과 같이 14개의 세부요소로 구성하였다.

<표 16> 데이터 공유 및 접근의 세부요소

RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
데이터 공유 및 접근								
데이터 이용 대상 및 방법 제시	○							
데이터 공유 계약 작성		○			○			
연구 참여자 기밀성 확보			○					
데이터 액세스(access) 플랫폼 제공		○	○	○	○	○	○	○
데이터 공개 제한 원칙 제시		○	○					
데이터 공개를 위한 체크리스트 제시						○		
데이터 재사용 체크리스트 제시							○	
기밀 유지 데이터 보호 방법 안내			○		○	○		
기밀 데이터의 접근 권한 부여			○	○	○	○		
기관 검토 위원회를 통한 데이터 공개 위험 평가			○					
저장 장치, 위치 및 액세스 계정 문서화				○				
데이터의 사용 및 재사용에 영향을 미치는 법률 및 정책 고려 사항 제시		○		○			○	

<표 17> 데이터 보존의 세부요소

RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
데이터 보존								
디지털 보존 정책 및 계획 수립			○					
보존을 위한 재난 계획 수립			○					
보존할 데이터 선택 기준 수립	○			○	○			
연구 데이터 기탁을 위한 도구 제시	○				○			
연구 데이터 기탁 동의서 제시	○				○			
연구 데이터 기탁 가이드라인 제공					○			
데이터 보존을 위한 파일 형식 제시		○						
파일 마이그레이션 제공		○						
기관 내 데이터 리포지토리 안내	○				○	○	○	○
신뢰할 수 있는 외부 디지털 리포지토리 안내	○		○			○	○	○
외부 리포지토리의 적합성 평가 기준 제시						○		
데이터 처분(이관) 안내					○	○		
연구 데이터의 보유 및 폐기 권장사항 제시							○	
데이터셋 보존 모범 사례 제시				○				

4.1.7 데이터 인용

‘데이터 인용’과 관련하여 8개 기관의 내용을 분석한 결과 22개의 세부요소가 도출되었다. 22개 중 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 인용’은 다음의 <표 18>과 같이 6의 세부요소로 구성하였다.

4.1.8 데이터 관리 교육

‘데이터 관리 교육’과 관련하여 8개 기관의 내용을 분석한 결과 16개의 세부요소가 도출되었다. 이러한 16개 중 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 관리 교육’은 다음의 <표 19>와 같이 6개의 세부요소로 구성하였다.

4.1.9 데이터 지적재산권

‘데이터 지적재산권’과 관련하여 8개 기관의 내용을 분석한 결과 17개의 세부요소가 도출되었다. 이러한 17개 중 유사하거나 동일하다고 판단되는 요소를 합한 결과 ‘데이터 지적재산권’은 다음의 <표 20>과 같이 7개의 세부요소로 구성하였다.

4.2 도출된 세부요소의 중요도

본 절에서는 4장 1절에서 도출된 세부요소에 대하여 설문 조사로 중요도를 파악하고자 한다.

4.2.1 세부요소의 중요도 조사

설문 조사의 목적은 기 도출된 Geoscience 분

<표 18> 데이터 인용의 세부요소

데이터 인용 \ RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
데이터 인용 원칙 제시	○				○			○
데이터 인용 구성요소 안내	○	○	○	○				
DOI 안내 및 서비스 제공	○	○			○	○		○
데이터 인용을 위한 도구 및 서비스 제시	○						○	
데이터 인용 예시 및 가이드라인 제시			○	○	○	○	○	○
데이터 인용 표준 및 모범 사례 제시					○	○		○

<표 19> 데이터 관리 교육의 세부요소

데이터 관리 교육 \ RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
교육 가이드와 헬프데스크 운영	○							
이용자별 교육 제공	○		○				○	
컨설팅 서비스 제공	○							
데이터 관리를 위한 오프라인 및 온라인 교육 제공		○	○	○				
교육 자료 제공		○				○	○	
해당 학문분야(도메인) 교육 제공							○	○

〈표 20〉 데이터 지적재산권의 세부요소

RDMS	DCC	ANDS	ICPSR	DataONE	BGS	USGS	CSIRO	IEDA
데이터 지적재산권								
데이터에 대한 지적 재산권 및 라이선스 적용 방법 안내		○	○	○	○	○		
해당 라이선스의 호환성 여부 확인 안내		○						
데이터 이용 및 배포 권한 부여 안내		○	○			○		
재사용하는 데이터에 대한 고려 사항 제시		○						
연구 데이터 라이선싱과 저작권을 위한 FAQ 제시		○						
저작권 자료의 상업적 및 비 상업적 이용 안내					○			
데이터 라이선스 계약서 제공					○		○	

야 RDM 기능별 세부요소에 대한 중요도를 파악하기 위함이다. 설문 조사는 '연구데이터'에 대하여 이해도가 높은 대학도서관 사서와 정부출연 연구소의 연구원 20명을 대상으로 하였다.

설문지는 크게 RDM 기능 9개의 중요도와 각 기능별 세부요소 중요도 2개의 섹션으로 구분하였고 중요도를 파악하기 위해 7점 척도를 사용하였다.

다음은 설문 조사의 개요를 나타낸 것이다.

- 조사 목적: Geoscience 분야 연구데이터 관리 및 서비스 모델 검증
- 조사 대상: 대학도서관 사서 및 정부출연 연구소 연구원
- 조사 일정: 2019년 11월 11일 ~ 2019년 11월 17일(7일간)
- 조사 방법: 웹 설문조사

#### 4.2.2 조사결과 분석

모든 설문 자료는 평균과 표준편차로 요약하였다. 본 연구에서 수집된 그룹의 크기가 20명으로 소표본(small sample)이기에 이표본 독립 T-검정 기법 사용을 위해 자료의 정규성 가정에

대해 검토하였다. 정규성 검토 결과 각 그룹이 소표본이지만 정규성 가정을 만족하여 모수적 기법인 T-검정을 사용할 수 있다. 본 분석을 위해 통계프로그램 R을 이용하였으며, 통계적 유의성을 위해 유의수준을 5%로 설정하였다. 즉, p-value(유의확률)가 0.05 이하인 경우 통계적으로 유의한 차이가 있는 것으로 판단하였다.

본 연구의 목적인 RDM 기능별 세부 요소에 대해 각 문항별 중요도를 살펴보기 위해 중요도 지수를 개발하고 이를 통해 각 세부요소별 중요도를 산정하여 우선순위를 선정하였다.

#### 4.3 기능별 세부요소 중요도

본 연구에선 기능별 각 세부 요소의 중요도를 살펴보기 위해 중요도 지수를 사용하였으며, 중요도 지수는 다음과 같은 식으로 정의하였다.

$$\text{중요도 지수} = \frac{\text{기능중요도} \times \text{세부요소중요도}}{49^*}$$

〈수식 1〉 중요도 지수

\* 49 = 핵심기능중요도의 최댓값 × 세부요소중요도의 최댓값



먼저 중요도는 9개의 기능 중요도와 각 기능에서 세부요소의 중요도를 곱하였기 때문에, 전체 세부 요소에서 가장 중요하게 처리해야 할 우선순위가 도출되었다. 즉, 중요도를 계산 한다는

것은 우선순위를 산정한다는 것을 의미한다. 다음의 <표 21>은 Geoscience 분야 연구데이터 관리 기능별 세부요소 중요도를 우선순위로 나열한 것이다.

<표 21> Geoscience 연구데이터 관리 기능별 세부요소

기능	순위	세부요소	중요도
DMP 작성지원	1	DMP 작성을 위한 템플릿 및 가이드 제시	0.855
	2	DMP 작성을 위한 샘플 및 체크리스트 제시	0.810
	3	DMP 작성 도구 제시	0.791
	4	DMP 구성요소 안내	0.771
	5	DMP 작성을 위한 교육과 컨설팅 제공	0.758
	6	DMP 평가	0.667
데이터 파일 정리	1	데이터 표준 제시	0.822
	2	파일 네이밍 규칙 및 모범사례 제시	0.766
	2	데이터 모델 정의	0.766
	4	권장하는 파일 포맷 제공	0.753
	5	데이터 및 파일 구조 제시	0.740
	6	데이터 버전 관리 방법 제공	0.734
	7	데이터 수집, 공개, 보존 시 파일 포맷에 관한 모범사례 제공	0.715
데이터 기술	1	메타데이터 표준 및 스키마 제공	0.862
	2	데이터와 메타데이터의 영구 연결을 지원하는 영구 식별자 제시	0.856
	3	메타데이터 생성	0.849
	4	메타데이터 공개	0.829
	5	메타데이터를 위한 시스템 설정	0.803
	6	메타데이터 수집, 업데이트 및 유지관리	0.796
	7	메타데이터 작성 모범사례 제시	0.790
	8	메타데이터 유효성 검증 도구 안내	0.770
	8	메타데이터 검토 체크리스트 제시	0.770
	10	메타데이터 레코드 작성 도구 안내	0.763
데이터 저장	1	저장 매체의 안정성 보장	0.822
	2	연구데이터 저장을 위한 스토리지 제공	0.810
	3	데이터 백업시 무결성 및 접근성 보장	0.784
	4	연구데이터 스토리지 할당 및 관리절차 개발	0.728
	4	프로젝트 초기에 데이터 저장 계획 수립	0.728
	6	정전시 백업 사이트에서 데이터 다운로드 서비스 제공	0.722
	7	저장할 연구데이터 사본의 개수와 동기화 방법 제시	0.690
	8	데이터 보안 모범 사례 제시	0.684
	9	클라우드 기반 서비스에 저장된 데이터 전송 시스템의 백업을 위한 사본 사이트 보유	0.678
	10	사본 저장 방법과 위치를 지정	0.665
	11	데이터 백업 모범 사례 제시 및 도구 안내	0.659
	12	스토리지 솔루션 비교 기준 제시	0.646

기능	순위	세부요소	중요도
데이터 공유 및 접근	1	기밀 데이터의 접근 권한 부여	0.823
	2	데이터 공개 제한 원칙 제시	0.797
	3	연구 참여자 기밀성 확보	0.784
	3	기밀 유지 데이터 보호 방법 안내	0.784
	5	데이터 액세스 플랫폼 제공	0.778
	5	저장 장치, 위치 및 액세스 계정 문서화	0.778
	7	데이터 공유 계약 작성	0.772
	7	데이터 공개를 위한 체크리스트 제시	0.772
	9	데이터 이용 대상 및 방법 제시	0.765
	10	데이터 재사용 체크리스트 제시	0.759
	11	기관 검토 위원회를 통한 데이터 공개 위험 평가	0.740
	12	데이터의 사용 및 재사용에 영향을 미치는 법률 및 정책 고려 사항 제시	0.733
데이터 보존	1	디지털 보존 정책 및 계획 수립	0.784
	1	연구데이터 기탁 동의서 제시	0.784
	3	연구데이터 기탁을 위한 도구 제시	0.778
	3	파일 마이그레이션 제공	0.778
	3	기관 내 데이터 리포지토리 안내	0.778
	6	데이터 보존을 위한 파일 형식 제시	0.771
	7	보존할 데이터 선택 기준 수립	0.765
	8	데이터 처분(이관) 안내	0.759
	9	보존을 위한 재난 계획 수립	0.752
	9	연구데이터 기탁 가이드라인 제공	0.752
	11	데이터셋 보존 모범 사례 제시	0.746
	12	연구데이터의 보유 및 폐기 권장사항 제시	0.739
	13	신뢰할 수 있는 외부 디지털 리포지토리 안내	0.726
13	외부 리포지토리의 적합성 평가 기준 제시	0.726	
데이터 인용	1	데이터 인용 원칙 제시	0.835
	2	DOI 안내 및 서비스 제공	0.829
	3	데이터 인용 구성요소 안내	0.791
	4	데이터 인용을 위한 도구 및 서비스 제시	0.778
	5	데이터 인용 예시 및 가이드라인 제시	0.759
	6	데이터 인용 표준 및 모범 사례 제시	0.753
데이터 관리 교육	1	교육 자료 제공	0.667
	2	교육 가이드와 헬프데스크 운영	0.640
	2	이용자별 교육 제공	0.640
	2	데이터 관리를 위한 오프라인 및 온라인 교육 제공	0.640
	5	컨설팅 서비스 제공	0.612
5	해당 학문분야(도메인) 교육 제공	0.612	
데이터 지적 재산권	1	데이터에 대한 지적 재산권 및 라이선스 적용 방법 안내	0.803
	2	데이터 라이선스 계약서 제공	0.796
	3	저작권 자료의 상업적 및 비 상업적 이용 안내	0.772
	4	데이터 이용 및 배포 권한 부여 안내	0.766
	5	해당 라이선스의 호환성 여부 확인 안내	0.753
	6	연구데이터 라이선싱과 저작권을 위한 FAQ 제시	0.747
	7	재사용하는 데이터에 대한 고려 사항 제시	0.741

## 5. 결 론

데이터에 대한 가치 증대와 더불어 연구 패러다임의 변화는 기존의 실험이나 이론 중심연구에서 데이터 기반 연구로 이동하고 있음을 보여주고 있다. 핵심 데이터의 확보와 데이터의 분석 및 활용의 중요성이 커짐에 따라 학문 분야별 데이터 관리는 필수가 되어 가고 추세이다. 해외 선진국에서는 이미 기관 및 대학에서 생성된 연구데이터의 탐색, 접근, 공유, 아카이빙 및 보존을 지원하기 위하여 연구데이터 관리 서비스를 계획하거나 구현하고 있지만 국내의 현황은 아직 걸음마 단계로서 법령에 연구데이터의 정의와 DMP 의무화에 대한 근거가 마련되어 있을 뿐이다. 또한 국내에서는 RDM 서비스 구축 및 운영 사례가 없어 RDM 기능별 세부요소를 도출하기 위해 영국의 DCC 등과 같은 국외 사례를 중심으로 연구를 진행하였다.

본 논문의 목적은 국내에서 Geoscience 분야 연구데이터 관리 시스템 개발에 적용할 수 있는 기능별 세부요소를 도출하는 것이다. 연구 목적 달성을 위해 수행한 연구 절차는 먼저, 4개의

RDM 서비스 기관과 4개의 Geoscience 분야 기관을 선정하여 분석을 수행하였다. 분석한 결과 80개의 RDM 기능별 세부 요소가 도출되었다. 마지막으로 국내 정보유통 분야의 연구자와 대학도서관 사서 등을 대상으로 Geoscience 분야 RDM 기능별 세부요소의 중요도를 파악하였다. 연구 결과 80개의 RDM 기능별 세부 요소가 도출되었으며 해당 요소는 중요도 순으로 4장에 제시되었다.

본 연구는 국내의 연구 기관 및 대학도서관 등에서 Geoscience 분야 RDM 서비스 구축 및 운영 시 가이드라인으로 활용될 수 있을 것이다. 물론, 이번 연구가 Geoscience라는 특정 분야를 중심으로 수행되었지만, RDM 기능별 세부요소를 도출하려고 참조한 사례 조사의 일부가 일반적인 RDM 서비스를 지원하고 있기 때문에 Geoscience 분야가 아니더라도 충분히 활용될 수 있을 것이라 판단된다. 마지막으로, 연구를 진행하기 위하여 활용된 연구 방법이 타 학문 분야 또는 커뮤니티를 위한 RDM 서비스를 연구하고자 할 때 유용한 참고자료가 되길 기대해 본다.

## 참 고 문 헌

- [1] 김성훈, 오삼균. 2018. 연구데이터 관리서비스의 구현 시 고려사항에 관한 연구. 『정보관리학회지』, 35(2): 141-165.
- [2] 김주섭, 김선태, 최상기. 2019. 연구 데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건. 『한국문헌정보학회지』, 53(3): 317-344.
- [3] 김지현. 2014. 대학도서관의 연구데이터관리서비스에 관한 연구: 미국 연구중심대학도서관을 중심으로. 『한국비블리아학회지』, 25(3): 165-189.

- [4] 신은정. 2018. 오픈사이언스시대 연구데이터의 공유·활용 동향과 시사점. 과학기술정책연구원, 2018년 6월 25일. [online] [cited 2019. 9. 2.]  
<[http://www.krnet.or.kr/board/data/dprogram/2295/C3-1\\_%BD%C5%C0%BA%C1%A4.pdf](http://www.krnet.or.kr/board/data/dprogram/2295/C3-1_%BD%C5%C0%BA%C1%A4.pdf)>
- [5] 심원식. 2015. 국가 차원의 연구데이터 관리체계 구축을 위한 로드맵 제안: 영국 사례 분석을 중심으로. 『한국문헌정보학회지』, 49(4): 355-378.
- [6] 심원식. 2016. 미국 대학도서관의 연구데이터 지원 서비스 사례 연구. 『한국문헌정보학회지』, 50(4): 311-332.
- [7] 이상환. 2019. DMP(Data Management Plan), 연구데이터 공유와 활용의 시작. 『'19년 한국전문도서관협의회-국가정책정보협의회 공동 학술세미나』, 2019년 6월 27일. 여수: 디오션리조트 호텔.
- [8] 한종규. 2019. 연구데이터 리퍼지토리 구축 경험과 과제(지질자원분야). 『'19년 한국전문도서관협의회-국가정책정보협의회 공동 학술세미나』, 2019년 6월 27일. 여수: 디오션리조트 호텔.
- [9] ANDS. 2019. ANDS About us. [online] [cited 2019. 7. 2.]  
<<https://www.andis.org.au/about-us>>
- [10] BGS. 2019. Data management. [online] [cited 2019. 6. 21.]  
<<https://www.bgs.ac.uk/services/NGDC/management/home.html>>
- [11] BGS. 2019. What is the British Geological Survey? [online] [cited 2019. 6. 21.]  
<<https://www.bgs.ac.uk/about/whoWeAre.html>>
- [12] Borghi, John, Abrams, Stephen, Lowenberg, Daniella, Simms, Stephanie and Chodacki, John. 2018. Support Your Data: A Research Data Management Guide for Researchers. Research Ideas and Outcomes. 4. e26439. 10.3897/rio.4.e26439.
- [13] Cox, Andrew M.; Verbaan, Eddy. 2018. EXPLORING RESEARCH DATA MANAGEMENT. London: Facet Publishing.
- [14] CSIRO. 2019. CSIRO Library Guides. [cited 2019. 6. 4.]  
<<http://libguides.csiro.au/c.php?g=586769&p=4053590>>
- [15] Eindhoven University of Technology. 2019. What is research data management?. [online] [cited 2019. 4. 20.]  
<<https://www.tue.nl/en/our-university/library/education-research-support/scientific-publishing/data-coach/general-terms-and-background/what-is-research-data-management/#top>>
- [16] Faundeen, J. L. et al. 2013. The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013-1265. Available online:  
<<http://dx.doi.org/10.3133/ofr20131265>>
- [17] Fearon, David Jr., Betsy Gunia, Sherry Lake, Barbara E. Pralle and Andrew L. Sallans.

2013. Research Data Management Services, SPEC Kit 334. Washington, DC: Association of Research Libraries. Available online: <<https://doi.org/10.29242/spec.334>>
- [18] Garry R. Baker. n.d. National Geoscience Data Centre (NGDC) 'Building an open national AGS data store'. [online] [cited 2019. 6. 7.]  
<[https://forumcourt.sharepoint.com/Ags/Events/Geotechnical%202017/4%20Garry%20Baker%20AGS\\_presentation\\_Aug17\\_v2.pdf](https://forumcourt.sharepoint.com/Ags/Events/Geotechnical%202017/4%20Garry%20Baker%20AGS_presentation_Aug17_v2.pdf)>
- [19] Guo, Huadong. 2017. "Big Earth data: A new frontier in Earth and information sciences." Big Earth Data 1.1-2: 4-20.
- [20] IEDA. 2019. Data Management Overview. [online] [cited 2019. 8. 2.]  
<<https://www.iedadata.org/help/data-management-overview/>>
- [21] Inter-university Consortium for Political and Social Research (ICPSR). 2012. About ICPSR. [online] [cited 2019. 6. 30.] <<https://www.icpsr.umich.edu/icpsrweb/content/about/>>
- [22] Jones, S., Pryor, G. and Whyte, A. 2013. 'How to Develop Research Data Management Services - a guide for HEIs'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <<http://www.dcc.ac.uk/resources/how-guides>>
- [23] Kraker, P., Leony, D., Reinhardt, W. and Beham, G. 2011. 'The case for an open science in technology enhanced learning', Int. J. Technology Enhanced Learning, 3(6): 643-654.
- [24] OECD. 2015. "Making Open Science a Reality." OECD Science, Technology and Industry Policy Papers. No. 25, Paris: OECD Publishing. [online] [cited 2019. 3. 8.]  
<<https://doi.org/10.1787/5jrs2f963zs1-en>>
- [25] Perrier, L. and Barnes, L. 2018. Developing research data management services and support for researchers: A mixed methods study. Partnership. The Canadian Journal of Library and Information Practice and Research, 13(1), 1-23.  
<<http://dx.doi.org/10.21083/partnership.v12i2.4115>>
- [26] Research Data Oxford. 2019. About RDM. [online] [cited 2019. 4. 19.]  
<<http://researchdata.ox.ac.uk/home/introduction-to-rdm/>>
- [27] Strasser et al. [n.d.]. DataONE Best Practices Primer. [online] [cited 2019. 8. 19.]  
<[https://www.dataone.org/sites/all/documents/DataONE\\_BP\\_Primer\\_020212.pdf](https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf)>
- [28] University College London. 2019. What is Research Data Management?. [online] [cited 2019. 4. 19.]  
<<https://www.ucl.ac.uk/library/research-support/research-data-management/best-practices>>
- [29] University of Pittsburgh. 2019. Research Data Management @ Pitt. [online] [cited 2019. 4. 19.] <<https://pitt.libguides.com/managedata>>

- [30] Whyte, A. and Tedds, J. 2011. 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available online: <<http://www.dcc.ac.uk/resources/briefing-papers>>
- [31] Yu, H. H. 2017. The role of academic libraries in research data service (RDS) provision. *The Electronic Library*, 35(4): 783-797.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kim, S. and Oh, S. G. 2018. Key Factors in the Implementation of Research Data Management Services. *Journal of the Korean Society for information Management*, 35(2): 141-165.
- [2] Kim, J. S., Kim, S. T. and Choi, S. K. 2019. The Functional Requirements of Core Elements for Research Data Management and Service. *Journal of the Korean Society for Library and Information Science*, 53(3): 317-344.
- [3] Kim, Jihyun. 2014. A Study on Research Data Management Services of Research University Libraries in the U.S. *Journal of the Korean BIBLIA Society for library and Information Science*, 25(3): 165-189.
- [4] Shin, Eun-jung. 2018. Trends and Implications of Sharing and Utilizing Research Data in the Open Science Era. Science and Technology Policy Institute. 25 June 2018. [online] [cited 2019. 9. 2.] <[http://www.krnet.or.kr/board/data/dprogram/2295/C3-1\\_%BD%C5%C0%BA%C1%A4.pdf](http://www.krnet.or.kr/board/data/dprogram/2295/C3-1_%BD%C5%C0%BA%C1%A4.pdf)>
- [5] Shim, Wonsik. 2015. Developing a Roadmap for National Research Data Management Governance: Based on the Analysis of United Kingdom's Case. *Journal of the Korean Society for Library and Information Science*, 49(4): 355-378.
- [6] Shim, Wonsik. 2016. A Case Study of U.S. Academic Libraries' Research Data Support Services. *Journal of the Korean Society for Library and Information Science*, 50(4): 311-332.
- [7] Lee, Sang Hwan. 2019. DMP (Data Management Plan) starts research data sharing and utilization. *19 Korea Special Library Association-National Association for Policy Information Joint Academic Seminar*, June 27, 2019. Yeosu: The Ocean Resort Hotel.
- [8] Han, Jong Kyu. 2019. DMP Research Data Repository Building Experience and Challenges (Geoscience). *19 Korea Special Library Association-National Association for Policy Information Joint Academic Seminar*, June 27, 2019. Yeosu: The Ocean Resort Hotel.