

# 연구 데이터 관리를 위한 데이터 라이프 사이클 제안

## Data Life Cycle Proposal for Research Data Management

김 주 섭 (Juseop Kim)\*

김 선 태 (Suntae Kim)\*\*

전 예 린 (Yerin Jeon)\*\*\*

### 목 차

- |                                 |                                    |
|---------------------------------|------------------------------------|
| 1. 서 론                          | 4. 구성 요소별 세부 내용 비교                 |
| 2. 이론적 배경                       | 5. 연구 데이터 관리를 위한 데이터 라이프 사이클 모델 제안 |
| 3. 연구 데이터 관리를 위한 데이터 라이프 사이클 분석 | 6. 결 론                             |

### 초 록

해외에서는 1990년대부터 데이터의 보존과 큐레이션을 위하여 데이터 라이프 사이클을 개발하였지만, 국내에서는 연구가 상당히 미흡한 실정이다. 본 연구에서는 연구 데이터 관리를 위한 데이터 라이프 사이클을 제안하기 위하여 DCC, ICPSR, IWGDD, DataONE, USGS 그리고 UKDA에서 개발된 데이터 라이프 사이클을 분석하였다. 분석 결과 공통적으로 도출된 구성 요소는 '계획', '생성 및 수집', '프로세스', '보존', '이관 및 폐기', '접근 및 이용', '기술', '보장' 그리고 '백업 및 보안' 등 9개이다. 또한 9개의 구성 요소를 단계별로 세분화하여 해당 단계에서 수행되어야 할 세부 내용을 기술하였다. 향후 국내에서 연구 데이터 관리를 위한 데이터 라이프 사이클을 개발할 때 본 연구의 내용을 적용할 수 있을 것으로 기대한다.

### ABSTRACT

Although overseas countries have already developed data life cycle for the preservation and curation of data since the 1990s, the research in Korea has been very insufficient. In this study, we analyzed the data life cycles developed in DCC, ICPSR, IWGDD, DataONE, USGS and UKDA to propose data life cycle for efficient management of research data. As a result of the analysis, the common components derived are 'Plan', 'Create & Collect', 'Process', 'Preserve', 'Dispose', 'Access & Use', 'Describe', 'Assure' and 'Backup & Secure'. In addition, the nine components were subdivided into stages to describe the details to be carried out at that stage. It is expected that the content of this study will be applicable in the future development of data life cycle for research data management in Korea.

키워드: 연구 데이터, 데이터 라이프 사이클, 보존, 큐레이션, 오픈 사이언스

Research Data, Data Life Cycle, Preservation, Curation, Open Science

\* 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 석사과정(niry96@gmail.com / ISNI 0000 0004 7936 5183) (공동저자)

논문접수일자: 2019년 10월 22일 최초심사일자: 2019년 11월 15일 게재확정일자: 2019년 11월 22일

한국문헌정보학회지, 53(4): 309-340, 2019. <http://dx.doi.org/10.4275/KSLIS.2019.53.4.309>

## 1. 서론

### 1.1 연구의 배경 및 목적

최근, 국가의 세금으로 생산된 연구 결과물이 국가 자산으로써 재사용될 수 있도록 체계적으로 관리 및 보존되어야 한다는 인식이 전세계적으로 확산되고 있다(김주섭, 김선태, 최상기 2019). 이와 같은 인식으로 자주 인용되는 있는 오픈 사이언스(Open Science)는 과학 연구(출판, 데이터, 샘플, 그리고 소프트웨어)를 학회, 일반인 또는 전문가가 쉽게 접근하도록 만드는 운동이다(Wikipedia 2019). 또한 오픈 사이언스는 연구 생산, 접근 및 활용 방식을 변화시키고 있으며 새롭고 다양한 과학적 교류 및 협력 방식을 가능케 하고 있다(European University Association 2019). 특히, 기존 문헌 연구에서 데이터 연구로 이동하는 연구 패러다임의 변화는 이러한 오픈 사이언스를 가능케 하는 동인 중 하나로 짐작될 수 있다. 2011년에 오픈 사이언스의 비전으로 오픈 액세스(Open Access), 오픈 데이터(Open Data), 오픈 소스(Open Source) 그리고 오픈 방법론(Open Methodology) 등 4가지 도구가 제시되었는데 이 중 오픈 데이터는 원시 데이터를 출판하기 위한 방법으로 이러한 데이터를 학술 영역에서 '연구 데이터'라고 명명하고 있다(Peter et al. 2011). 최근 이러한 연구 데이터를 보다 효과적으로 관리하기 위하여 연구 데이터 관리(Research Data Management, 이하 RDM)에 대한 관심이 높아지고 있다. 여기에 맞추어 많은 기관 및 대학도서관에서는 해당 기관 및 대학에서 생성된 연구 데이터의 탐색, 접근, 공유, 아카이빙

및 보존을 지원하는 연구 데이터 관리 서비스를 계획하거나 구현하고 있는 추세이다. 연구 데이터는 많은 시간과 돈을 투자하여 만든 중요한 자원으로서 연구의 투명성 강화, 데이터 접근, 시간 절약, 데이터 손실 위험 감소, 재사용과 공유 촉진 그리고 인용의 증가 등을 위해 적절히 관리되어야 한다(University of Leeds 2019). 또한 미국의 의학 도서관 네트워크인 NNLM에서도 데이터를 관리해야 하는 이유에 대하여 연구 영향력 향상, 시간 절약, 데이터 보존, 데이터 무결성 유지, 펀딩 기관의 요구 조건 충족, 새로운 발견 홍보 그리고 오픈 액세스 지원 등을 제시하고 있다(NNLM 2019).

연구 데이터는 대부분 프로젝트 중에 생산되며 연구 데이터의 수명이 프로젝트보다 수명이 길기 때문에 프로젝트가 완료된 후에도 데이터를 관리하는 방법을 계획해야 한다. 즉, 연구 데이터의 생성, 구성, 보존, 공유 및 재사용하는 방법에 대하여 생각해야 하는데 보통 이러한 과정을 연구 데이터 라이프 사이클(research data life cycle)이라 부른다. 이러한 연구 데이터 라이프 사이클은 효과적인 연구를 설계할 수 있으며 기존 데이터에 어떻게 데이터가 추가되는지, 데이터의 수집 및 처리 그리고 재사용할 데이터를 언제 어디에 기탁할지를 결정하는 데 도움을 줄 수 있다(University of Leeds 2019).

해외에서는 이미 1990년대부터 데이터 큐레이션 지원하기 위한 목적으로 데이터 라이프 사이클이 연구 기관을 중심으로 제시되고 있는 실정이지만(Louise et al. 2014), 국내의 경우 최근 몇 년 전부터 연구 데이터에 관한 연구가 진행되었으며, 현재는 데이터 관리 계획(DMP)의 작성 의무화 등이 2019년 9월 1일자로 시행

되었을 뿐이다(국가연구개발사업의 관리 등에 관한 규정 2019).

본 연구의 목적은 연구 데이터 관리를 위하여 해외에서 개발된 기관별 데이터 라이프 사이클을 비교 분석하여 국내의 대학 도서관 및 연구 기관에서 적용할 수 있는 연구 데이터 라이프 사이클 모델을 제시하는 것을 목적으로 하였다. 이를 위하여 영국의 데이터 큐레이션 센터에서 개발된 'DCC 큐레이션 라이프 사이클(Curation life cycle)', 사회과학 데이터 센터인 'ICPSR의 데이터 라이프 사이클', 디지털 데이터에 관한 기관 간 워킹 그룹인 'IWGDD 데이터 라이프 사이클', 과학 데이터 아카이빙을 위한 네트워크인 'DataONE 데이터 라이프 사이클', 지질구조와 광물자원을 평가하기 위한 연구 기관인 'USGS 데이터 라이프 사이클' 그리고 사회과학 및 인구와 관련한 디지털 리포지토리인 'UK Data Archive 데이터 라이프 사이클' 등 6개의 데이터 라이프 사이클을 분석 대상으로 삼았다.

## 1.2 연구 방법과 범위

연구 데이터 관리를 위한 데이터 라이프 사이클 제안을 위하여 본 연구에서 분석한 데이터 라이프 사이클 목록은 다음과 같다.

- DCC(Digital Curation Centre)
- ICPSR(Inter-university Consortium for Political and Social Research)
- IWGDD(Interagency Working Group on Digital Data)
- DataONE(Data Observation Network

for Earth)

- USGS(U.S. Geological Survey)
- UK Data Archive

우선, 본 연구에서는 연구 데이터 관리를 위해 개발된 모든 데이터 라이프 사이클을 분석하는 데 한계가 있기 때문에 위에서 제시하는 6개의 데이터 라이프 사이클을 선택하여 연구의 범위를 제안하였다. DCC는 영국의 연구 데이터 큐레이션을 위한 정책을 제시하는 기관으로서 2008년에 데이터 라이프 사이클을 개발하였으며, ICPSR의 경우, 전통적인 사회과학 분야 데이터 아카이브로서 사회과학 분야에 특화된 데이터 라이프 사이클을 제시하여 본 연구 대상으로 선택하였다. 다음으로 IWGDD는 과학, 기술, 공학 분야의 연구, 개발, 교육을 위한 디지털 데이터의 신뢰성 있는 보존과 효과적인 접근을 보장하기 위해 설립된 워킹 그룹으로 여러 학문분야의 특성에 맞게 데이터 라이프 사이클을 제시하였다. DataONE은 지구 관측에 대한 다중 규모, 다 분야 및 다국적 과학 데이터의 보존, 접근, 이용 및 재사용을 보장하기 위해 설립된 네트워크이며 성공적인 인프라로 평가받고 있어 선택하였다. USGS는 미국의 지질 분야 조사 연구 기관으로서 데이터가 잘 기술되고 보존되고 접근되며 재사용될 수 있도록 과학 데이터 라이프 사이클 모델(Science Data life cycle Model, SDLM)을 개발하여 이번 연구 범위에 포함하였다. 마지막으로 영국 데이터 아카이브(UK Data Archive)는 사회, 경제 및 인구 분야의 신뢰할 수 있는 디지털 리포지토리로서 데이터 관리 분야의 국제적인 모범 사례를 제시하고 있어 선택하였다.

본 연구에서 제시된 목적을 달성하기 위하여 먼저, 선택된 6개 기관의 데이터 라이프 사이클의 구성 요소를 중심으로 세부 내용을 분석하였다. 다음으로 6개 기관에서 공통적으로 도출된 구성 요소를 중심으로 기관별로 해당 내용을 비교 분석하여 최종적으로 데이터 라이프 사이클을 제안하였다. 다음의 <그림 1>은 본 연구의 목적인 연구 데이터 관리를 위해 데이터 라이프 사이클을 제안하기 위한 연구 절차를 나타낸 것이다.

### 1.3 연구의 제한점

연구 방법에서 나타난 바와 같이 국내에서는 연구 데이터 관리를 위한 데이터 라이프 사이클에 대하여 연구가 전무한 실정이다. 국내의 경우, 몇 년 전부터 도서관 및 연구 기관에서 연구 데이터 관리를 위한 연구가 본격화되었고 현재 「국가연구개발사업의 관리 등에 관한 규정」이라는 법령을 통해 연구 데이터 정의와 데이터 관리 계획 도입 및 DMP 정보 등록 근거가 제시되었다. 이에 반해 연구 데이터를 관리하기 위하여 이미 해외의 대학 도서관 및 연구 기관에서는 연구 주기를 파악하여 데이터 라이프 사이클을 제시하고 있으며 제시된 라이프 사

이클 단계별로 연구자에게 필요한 모범 사례를 구체적으로 제시하고 있는 실정이다(DataONE Public Participation in Scientific Research Working Group 2013). 본 연구에서는 제시된 6개의 사례를 중심으로 문헌분석을 통하여 데이터 라이프 사이클의 구성 요소와 해당 요소별 세부 내용을 추출하였다. 물론, 본 연구가 국내에서 연구 데이터 관리를 위한 데이터 라이프 사이클을 제시한다는 점에서 의의를 가질 수 있지만, 데이터 라이프 사이클에 영향을 미칠 수 있는 요인 분석과 연구자의 인터뷰나 설문조사 등 실증적 검증이 이루어지지 않았다는 점에서 한계를 가진다. 또한, 향후, 도서관 또는 연구 기관에서 연구 데이터 관리를 위한 데이터 라이프 사이클을 제시할 때 본 연구가 기초적인 가이드라인으로 제시될 수 있을 것으로 기대하지만, 연구 초기 연구자와의 면담을 통한 요구조건 수렴 그리고 제시된 라이프 사이클에 대한 실증적인 검증 등이 추가적으로 필요할 것으로 판단된다.

## 2. 이론적 배경

이번 장에서는 데이터 라이프 사이클의 개요

1단계	연구 방향 및 범위 설정을 위한 브레인 스토밍
2단계	기관별 데이터 라이프 사이클 분석 - DCC 등
3단계	데이터 라이프 사이클의 공통 구성 요소 도출
4단계	도출된 공통 구성 요소의 세부 내용 비교 - DCC 등
5단계	연구 데이터 관리를 위한 데이터 라이프 사이클 제안

<그림 1> 데이터 라이프 사이클 모델 개발 절차

와 해당하는 선행 연구를 제시하였다. 1990년대와 2000년대 초반에 디지털 보존과 데이터 큐레이션을 지원하기 위한 목적으로 추진된 데이터 라이프 사이클은 데이터 공유 문화가 일상 연구 언어의 일부가 되면서 인기를 얻은 개념이다. 이러한 데이터 라이프 사이클은 일반적인 연구 사이클을 확장한 것이며 일반적으로 수행되는 데이터 관련 활동의 개요를 설명한 것이다. 여기에는 '탐색과 계획', '데이터 수집', '데이터 처리 및 분석', '출판 및 공유', '장기 관리' 그리고 데이터 재사용 등 6개의 데이터 관련 활동을 포함하고 있다(Louise et al. 2014). 또한, 연구 데이터 라이프 사이클 모델은 성공적인 데이터 큐레이션 및 보존을 보장하기 위해 연구 사이클에서 수행해야 할 단계를 설명하고 식별한다. 이러한 연구 데이터 라이프 사이클에는 데이터 생성, 데이터 처리, 데이터 분석 등이 포함된다(NTU Library 2019).

Pennock(2007)은 디지털 자료의 연속성을 보장하기 위해 라이프 사이클 접근 필요성을 논의하였는데, 디지털 자료는 본질적으로 만들어지는 순간부터 기술변화에 영향을 받기 쉬우므로 라이프 사이클 접근 방식을 통해 모든 필수 단계를 식별하고 계획해야만 필요한 조치를 올바른 순서로 구현할 수 있으며, 이를 통해 디지털 자료의 진위성, 신뢰성, 무결성 및 유용성을 유지하여 제작에 대한 투자를 극대화할 수 있다고 하였다.

C Jung 등(2014)은 데이터가 대부분의 과학 분야에서 중심적인 역할을 하고 있으며 최근 몇 년 동안 데이터의 양이 급격하게 증가하였고 데이터도 복잡해졌다고 제시하면서 과학별 데이터에 적용되는 방법과 기법은 커뮤니티

특정 요구에 초점을 맞추어 수정되어야 한다고 밝히고 있다. 이를 위해 저자는 사례 분석을 통하여 데이터 라이프 사이클의 효율성과 최적화에 대하여 논의하였다. 분석된 사례는 고에너지 물리학에서 시스템 생물학에 이르기까지 다양한 연구 분야에서 데이터 라이프 사이클의 효율성을 극대화하는 것을 목표로 한 독일 헬름홀츠 협회 프로젝트인 '대규모 데이터 관리 및 분석(LSDMA)'이다. 해당 프로젝트 내 데이터 라이프 사이클 랩(Data Life Cycle Labs, DCLC)의 데이터 전문가는 공동 연구 개발 분야의 커뮤니티와 긴밀히 협력하여 각 데이터 라이프 사이클을 최적화하며 데이터 서비스 통합 팀(DSIT)은 여러 DLCL에 공통적인 데이터 분석 툴과 서비스를 제공한다.

Tanja와 Matej(2015)는 『연구 데이터 워크플로우: 연구 데이터 라이프 사이클 모델에서 기관 솔루션』이라는 논문에서 현재 오스트리아 과학 아카데미에서 시행되고 있는 연구 데이터에 대한 제도적 워크플로우 모델을 제시했는데, 이 모델은 제도적, 국가적 맥락에서 예술과 인문학의 연구자들을 위한 서비스 단위 역할도 하는 아카데미의 신생 연구소인 ACDH-OEAW가 조정하였다. 추상적(연구) 데이터 라이프 사이클 모델에서 시작하여, 특정 제도적 설정에 대한 이해당사자와 시나리오를 논의하였고, 아카데미의 특정 상황을 촉진하는 워크플로우 모델을 상세하게 기술하였다.

Mohammed 등(2017)은 데이터를 지식으로 변환하려면 기대에 맞는 적절한 라이프 사이클을 선택해야 하며 이를 위해 관련 기준을 파악하고 각 기준에 대한 등급 시스템을 정의했다. 연구를 위해 저자는 Information pyramid, CRUD

life cycle, life cycle for Big Data, IBM life cycle, DataOne life cycle, Information life cycle, CIGREF, DDI life cycle, USGS life cycle, PII life cycle, Enterprise data life cycle 그리고 Hindawi life cycle 등 12개의 라이프 사이클을 분석하여 필수 단계를 제시하였다. 필요한 단계는 생성/수집, 분석, 저장 그리고 아카이빙이며, 나머지는 선택 사항이지만 데이터를 더 잘 제어하기 위해 중요하다고 제시하였다.

Kristi 등(2018)은 「데이터 관리 라이프 사이클」이라는 보고서를 통해 향후 연구를 위해 높은 수준의 우선순위 지정에 사용할 데이터 관리 로드맵을 제공하였다. 저자들은 데이터 관리 라이프 사이클을 개발하여 데이터를 구성하고, 시간에 따른 데이터의 성격과 가치를 특징짓고, 코로스 커팅(cross-cutting) 데이터 관리 이슈의 정책적 함의를 파악하였다. 특히, 이 연구는 라이프 사이클에서 데이터가 이동하는 7가지 단계와 라이프 사이클 단계에 영향을 미치는 이슈에 대하여 다음과 같이 제시하였다.

〈표 1〉 데이터 라이프 사이클과 영향을 미치는 이슈

데이터 라이프 사이클	데이터 라이프 사이클에 영향을 미치는 이슈
<ul style="list-style-type: none"> <li>• 수집</li> <li>• 프로세스</li> <li>• 저장 및 보안</li> <li>• 이용</li> <li>• 공유 및 커뮤니케이션</li> <li>• 아카이빙</li> <li>• 파기 또는 재사용(동시 단계)</li> </ul>	<ul style="list-style-type: none"> <li>• 목적 및 가치</li> <li>• 프라이버시</li> <li>• 데이터 소유권</li> <li>• 책임</li> <li>• 공공 인식</li> <li>• 보안</li> <li>• 표준 및 데이터 품질</li> </ul>

이러한 선행 연구를 분석한 결과, 해외에서

는 이미 1990년대부터 데이터의 보존과 큐레이션을 위해 개념적 모델부터 실제적 적용을 위한 모델이 개발되어 왔다. 이에 반해 국내에서는 데이터 관리를 위한 데이터 라이프 사이클에 대한 연구가 전무하여 현황을 분석하지 못하였다. 따라서 본 연구에서는 연구 데이터 관리를 위한 데이터 라이프 사이클 모델을 최초로 제시한다는 점에서 본 연구의 의의를 가지며, 또한 도출된 구성 요소에 한하여 세부 내용을 제시함으로써 실질적인 연구 데이터 관리를 위한 데이터 라이프 사이클 모델을 구현할 때 실무에 활용할 수 있을 것이다.

### 3. 연구 데이터 관리를 위한 데이터 라이프 사이클 분석

이번 장에서는 데이터 라이프 사이클을 제안하기 위하여 선택된 6개 기관의 데이터 라이프 사이클을 분석하였다. 다음의 〈표 2〉는 이번 장에서 분석할 데이터 라이프 사이클을 제시한 것으로, 해당 내용은 6개의 기관에서 제시한 데이터 라이프 사이클의 구성 요소를 정리한 것이다.

#### 3.1 DCC(Data Curation Centre)

DCC는 초기 개념화에서 폐기 또는 재사용 및 장기 보존을 위한 선택에 이르기까지 성공적인 큐레이션 및 보존을 가능하게 하고 디지털 자료 관리에 대한 라이프 사이클 접근 방식을 홍보하기 위해 노력하고 있다(Higgins 2008). 이러한 Higgins(2008)에 의해 제안된 DCC의

〈표 2〉 분석 대상 데이터 라이프 사이클

데이터 라이프 사이클	세부 내용	비고
DCC	<ul style="list-style-type: none"> <li>• 전체 라이프 사이클 활동(4개 요소)</li> <li>• 순차적 활동(8개 요소)</li> <li>• 비정기적 활동(3개 요소)</li> </ul>	15개 요소
ICPSR	<ul style="list-style-type: none"> <li>• 1단계: 제안서 개발 및 데이터 관리 계획(DMP)</li> <li>• 2단계: 프로젝트 시작</li> <li>• 3단계: 데이터 수집 및 파일 생성</li> <li>• 4단계: 데이터 분석</li> <li>• 5단계: 공유할 데이터 준비</li> <li>• 6단계: 데이터 기탁</li> </ul>	6단계
IWGDD	<ul style="list-style-type: none"> <li>• 계획, 생성, 보존, 기술 습득 및 구현, 이관 및 폐기, 문서화, 구성, 보호 및 접근</li> </ul>	8개 요소
DataONE	<ul style="list-style-type: none"> <li>• 계획, 수집, 보장, 기술, 보존, 탐색, 통합, 분석</li> </ul>	8개 요소
USGS	<ul style="list-style-type: none"> <li>• 계획, 수집, 프로세스, 분석, 보존, 출판/공유, 기술(메타데이터, 문서화, 품질 관리, 백업 및 보안)</li> </ul>	9개 요소
UKDA	<ul style="list-style-type: none"> <li>• 생성, 처리, 분석, 보존, 접근 권한 부여, 재사용</li> </ul>	6개 요소

큐레이션 라이프 사이클 모델은 초기 개념화에서 데이터를 성공적으로 큐레이션하고 보존하는데 필요한 단계를 제시하고 있으며 조직 또는 컨소시엄 내 활동을 계획하여 필요한 모든 단계를 올바른 순서대로 수행할 수 있도록 도와준다. 다음의 〈표 3〉은 DCC의 큐레이션 라이프 사이클 모델의 내용을 기술한 것이다.

DCC의 큐레이션 라이프 사이클은 3개의 활동, 15개의 하위 요소로 구성되어 있다. 3개의 활동은 전체 라이프 사이클 활동, 순차적 활동 그리고 비정기적 활동을 포함한다.

전체 라이프 사이클 활동은 관련 메타데이터를 할당하여 필요한 표현 정보를 수집하고 할당하며 디지털 자료의 큐레이션 라이프 사이클 동안 보존 계획을 제시한다. 또한 관련 커뮤니티 활동의 관찰을 통해 적합한 소프트웨어 개발에 참여하도록 하고 있다. 해당 하위 요소는 '기술 및 표현 정보', '보존 계획', '커뮤니티 관찰 및 참여' 그리고 '큐레이션 및 보존' 등 4가

지고 구성되었다.

순차적 활동은 데이터 생성을 계획하는 '개념화', 데이터를 생성하거나 리포지토리 또는 데이터 센터에서 데이터를 수집하고 메타데이터를 할당하는 '생성 및 입수', 데이터를 평가하고 장기 큐레이션 및 보존을 위한 선택 과정인 '평가 및 선택', 아카이브 및 리포지토리에 데이터를 전송하는 '수집', 데이터 관리 특성을 장기간 보존하고 보유하기 위한 조치인 '보존 활동', 관련 표준을 준수하는 안전한 방식으로 데이터를 저장하는 '저장', 이용자가 데이터에 접근할 수 있는지에 대한 '접근, 이용 및 재사용' 마지막으로 원본에서 새 데이터를 생성하는 '변환' 등을 포함하고 있다.

마지막 비정기적 활동은 '폐기', '재평가' 그리고 '마이그레이션'을 하위 요소로 제시하고 있다. '폐기'는 장기 큐레이션 및 보존을 위해 선택되지 않은 데이터를 폐기하는 과정을 '재평가'는 추가 평가 및 재-선택을 위한 검증 절차에 실패

〈표 3〉 DCC 큐레이션 라이프사이클 모델

활동	내용	
전체 라이프 사이클 활동	기술 및 표현 정보	<ul style="list-style-type: none"> <li>• 관리, 설명, 기술, 구조 및 보존 메타데이터를 할당하여 장기간에 걸쳐 설명 및 제어를 보장</li> <li>• 디지털 자료와 관련 메타데이터를 이해하고 제공하는 데 필요한 표현 정보를 수집하고 할당</li> </ul>
	보존 계획	<ul style="list-style-type: none"> <li>• 모든 큐레이션 라이프 사이클 활동의 관리와 운영에 대한 계획과 디지털 자료의 큐레이션 라이프 사이클 동안의 보존 계획</li> </ul>
	커뮤니티 관찰 및 참여	<ul style="list-style-type: none"> <li>• 커뮤니티 활동을 주시하고 공유 표준, 도구 및 적합한 소프트웨어 개발에 참여</li> </ul>
	큐레이션 및 보존	<ul style="list-style-type: none"> <li>• 라이프 사이클 동안 큐레이션 및 보존을 촉진하기 위해 계획된 관리 및 운영 작업 수행</li> </ul>
순차적 활동	개념화	<ul style="list-style-type: none"> <li>• 캡처 방법 및 스토리지 옵션을 포함하여 데이터 생성을 고안하고 계획</li> </ul>
	생성 및 입수	<ul style="list-style-type: none"> <li>• 관리, 설명, 구조 및 기술 메타데이터를 포함한 데이터 생성, 생성 시 보존 메타데이터가 추가될 수도 있음</li> <li>• 문서화된 수집 정책에 따라 데이터 작성자, 기타 아카이브, 리포지토리 또는 데이터 센터에서 데이터를 수집하고 필요한 경우 적절한 메타데이터를 할당</li> </ul>
	평가 및 선택	<ul style="list-style-type: none"> <li>• 문서화된 지침, 정책 또는 법적 요구 사항을 준수하여 데이터를 평가하고 장기 큐레이션 및 보존을 위해 선택</li> </ul>
	수집	<ul style="list-style-type: none"> <li>• 문서화된 지침, 정책 또는 법적 요구 사항을 준수하여 아카이브, 리포지토리, 데이터 센터 또는 기타 관리자에게 데이터를 전송</li> </ul>
	보존 활동	<ul style="list-style-type: none"> <li>• 데이터의 관리 특성을 장기간 보존하고 보유하기 위한 조치</li> <li>• 보존 활동은 데이터의 무결성을 유지하면서 진본, 신뢰성 및 가용성을 보장해야 함</li> <li>• 작업에는 데이터 정리, 유효성 검사, 보존 메타데이터 할당, 표현 정보 할당 및 허용 가능한 데이터 구조 또는 파일 형식 보장이 포함</li> </ul>
	저장	<ul style="list-style-type: none"> <li>• 관련 표준을 준수하는 안전한 방식으로 데이터 저장</li> </ul>
	접근, 이용 및 재사용	<ul style="list-style-type: none"> <li>• 지정된 이용자와 재사용자 모두가 매일 데이터에 액세스 할 수 있는지 확인</li> <li>• 이것은 공개적으로 이용 가능한 출판된 정보의 형태일 수 있음</li> <li>• 강력한 접근 제어 및 인증 절차가 적용될 수 있음</li> </ul>
비정기적인 활동	폐기	<ul style="list-style-type: none"> <li>• 문서화된 정책, 지침 또는 법적 요구 사항에 따라 장기 큐레이션 및 보존을 위해 선택되지 않은 데이터 폐기</li> <li>• 일반적으로 데이터는 다른 아카이브, 리포지토리, 데이터 센터 또는 기타 관리자에게 전송될 수 있음</li> <li>• 어떤 경우에는 데이터가 파괴되며 법적인 이유로 데이터의 특성상 안전한 폐기가 필요할 수 있음</li> </ul>
	재평가	<ul style="list-style-type: none"> <li>• 추가 평가 및 재-선택을 위한 검증 절차에 실패한 데이터 반환</li> </ul>
	마이그레이션	<ul style="list-style-type: none"> <li>• 데이터를 다른 형식으로 마이그레이션</li> <li>• 스토리지 환경에 따라 수행되거나 하드웨어 또는 소프트웨어 노후화로 인한 데이터의 내성을 보장하기 위해 수행될 수 있음</li> </ul>



한 데이터 반환을 말한다. ‘마이그레이션’은 하드웨어 또는 소프트웨어 노후화로 인한 데이터의 내성을 보장하기 위해 다른 형식으로 변환하는 과정이다.

### 3.2 ICPSR(Inter-University for Political and Social Research)

1962년에 설립된 ICPSR(Inter-University for Political and Social Research)은 750개가 넘는 학술 기관 및 연구 기관으로 구성된 국제 컨소시엄으로서 사회 과학 연구 커뮤니티를 위한 데이터 접근 큐레이션 및 분석 방법에 대한 리더십과 교육을 제공한다. 이러한 서비스를 제공하기 위해 ICPSR은 데이터 관리 및 큐레이션(Data Management & Curation) 서비스를 6가지 카테고리로 범주화하여 제시하고 있다. 6가지는 품질, 보존, 접근, 기밀 유지, 인용, 도구 및 서비스로 구성되어 있으며 이 중 ‘품질(Quality)’의 세부 내용인 데이터 아카이빙을 통해 데이터 라이프 사이클을 제공하고 있다(ICPSR 2019).

다음의 <표 4>는 ICPSR의 데이터 라이프 사이클을 단계별로 정리한 것이다.

1단계는 데이터 자원을 장기적으로 이용이 가능하도록 보장하기 위한 데이터 관리 계획(Data Management Plan, 이하 DMP)에 대한 작성을 포함하고 있다. 펀딩 기관이 지원 신청서에 데이터 공유 및 배포계획을 포함하도록 요구함에 따라 연구자는 해당 단계에서 데이터 기탁 및 장기 보존 계획을 수립해야 한다. 또한 DMP에는 13개의 추천 요소와 4개의 선택 요소로 구성되어 있다. 13개의 추천 요소에는 ‘데

이터 기술’, ‘현용 데이터’, ‘포맷’, ‘메타데이터’, ‘스토리지 및 백업’, ‘보안’, ‘책임’, ‘지적재산권’, ‘접근 및 공유’, ‘이용자’, ‘선택 및 보유 기간’, ‘아카이빙 및 보존’, ‘윤리 및 프라이버시’ 등이 포함되고 4개의 선택 요소에는 ‘예산’, ‘데이터 정리’, ‘품질 보장’, ‘법적 요구사항’ 등으로 구성되어 있다.

프로젝트 시작 단계인 2단계는 ‘고려해야 할 초기 질문’, ‘데이터 및 도큐멘테이션 통합’ 그리고 ‘사전 테스트 및 파일럿 연구의 일부로 데이터 입력 및 도큐멘테이션’ 등 세 가지 구성요소로 되어 있다. 첫 번째 ‘고려해야 할 초기 질문’에는 ‘데이터 및 파일 구조’, ‘네이밍 규칙’, ‘데이터 무결성’, ‘데이터셋 도큐멘테이션 준비’, ‘변수 구성’ 그리고 ‘프로젝트 도큐멘테이션’ 등을 포함한다. 두 번째는 ‘데이터 및 도큐멘테이션 통합’으로서 프로젝트 초기에 데이터 관리 및 도큐멘테이션 작업 처리 프로그램을 결정하는 것으로 ‘컴퓨터 지원 인터뷰’, ‘통합 소프트웨어 사용’ 그리고 ‘스프레드시트 및 데이터베이스’ 등을 포함하고 있다. 세 번째는 ‘테스트 및 파일럿 연구의 일부로 데이터 입력 및 도큐멘테이션’이다. 프로젝트 시작 시 테스트 및 파일럿 연구를 수행하면 해당 프로젝트의 비용을 추정할 수 있으며 또한, 데이터 입력 테스트 및 도큐멘테이션은 레코드 레이아웃, 네이밍 규칙 등에 대한 예상치 못한 잠재적인 문제를 발견할 수 있다.

3단계에서는 사회과학의 정량적, GIS(지리 정보 체계), 정성적, 기타 유형의 데이터 및 메타데이터 등에 대해 널리 수용되는 기준에 관해 설명하고 있다. 해당 단계에서는 데이터셋 무결성, 변수명, 레이블 및 그룹; 코딩; 누락 데

〈표 4〉 ICPSR의 Data Life Cycle

단계	내용	
1단계: 제안서 개발 및 데이터 관리 계획(DMP)	<ul style="list-style-type: none"> <li>• 제안서 개발</li> <li>• 데이터 자원의 장기 가용성을 보장하기 위한 DMP 작성</li> </ul>	
2단계: 프로젝트 시작	<ul style="list-style-type: none"> <li>• 고려해야 할 초기 질문</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 및 파일 구조</li> <li>- 네이밍 규칙</li> <li>- 데이터 무결성</li> <li>- 데이터셋 도큐멘테이션 준비</li> <li>- 변수 구성</li> <li>- 프로젝트 도큐멘테이션</li> </ul>
	<ul style="list-style-type: none"> <li>• 데이터 및 도큐멘테이션 통합</li> </ul>	<ul style="list-style-type: none"> <li>- 컴퓨터 지원 인터뷰</li> <li>- 통합 소프트웨어 사용</li> <li>- 스프레드시트 및 데이터베이스</li> </ul>
	<ul style="list-style-type: none"> <li>• 사전 테스트 및 파일럿 연구의 일부로 데이터 입력 및 도큐멘테이션</li> </ul>	
3단계: 데이터 수집 및 파일 생성	<ul style="list-style-type: none"> <li>• 정량적 데이터</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터셋 생성 및 무결성</li> <li>- 변수명</li> <li>- 변수 라벨</li> <li>- 변수 그룹</li> <li>- 코드 및 코드북</li> <li>- 누락 데이터</li> <li>- 누락 데이터 코드 선택</li> <li>- 결측값 대체 데이터(Imputed data)</li> <li>- 지리 식별자 및 지리 공간 데이터</li> </ul>
	<ul style="list-style-type: none"> <li>• 정성적 데이터</li> </ul>	<ul style="list-style-type: none"> <li>- 정성적 데이터의 유형</li> <li>- 정성적 데이터의 기밀성</li> <li>- 정성적 데이터를 위한 도큐멘테이션</li> </ul>
	<ul style="list-style-type: none"> <li>• 기타 데이터 유형</li> </ul>	
	<ul style="list-style-type: none"> <li>• 메타데이터 생성 모범 사례</li> </ul>	<ul style="list-style-type: none"> <li>- XML</li> <li>- 데이터 도큐멘테이션 이니셔티브(DDI)</li> <li>- DDI 제작 옵션</li> <li>- DDI 메타데이터 기탁</li> <li>- 중요한 메타데이터 요소</li> </ul>
4단계: 데이터 분석	<ul style="list-style-type: none"> <li>• 마스터 데이터셋과 작업 파일(work files)</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 및 도큐멘테이션 버전 관리</li> <li>- 원시 데이터와 통계 시스템 파일</li> <li>- 파일 구조</li> <li>- 데이터 백업</li> </ul>
5단계: 공유할 데이터 준비	<ul style="list-style-type: none"> <li>• 응답자 기밀성</li> </ul>	<ul style="list-style-type: none"> <li>- 공개 위험 제한 원칙</li> <li>- 기밀 보호 관행</li> <li>- 이용이 제한된 데이터 수집</li> <li>- 데이터 Enclave</li> </ul>
6단계: 데이터 기탁	<ul style="list-style-type: none"> <li>• 파일 형식</li> </ul>	<ul style="list-style-type: none"> <li>- 소프트웨어별 시스템 파일</li> <li>- 휴대용 소프트웨어별 파일</li> <li>- ASCII 데이터 및 설정 파일</li> <li>- 온라인 분석 가능 파일</li> <li>- 기타 파일 형식</li> </ul>
	<ul style="list-style-type: none"> <li>• 기존 또는 2 차 데이터 분석에서 파일 아카이브</li> </ul>	<ul style="list-style-type: none"> <li>- 링크드 데이터</li> </ul>

이터에 대한 처리 기준과 생성된 변수의 문서화 등 관련된 도큐멘테이션 요소를 포함한 DDI 표준을 사용하도록 제안하고 있다. 특히, 중요한 메타데이터 요소는 19개 항목을 제시하고 있는데 해당 항목은 더블린 코어 메타 데이터 이니셔티브(DCMI) 요소 세트를 참조하고 있다.

4단계는 '데이터 분석'으로서 연구 질문을 조사하기 위하여 데이터 파일을 적극적으로 연구할 때 분석 단계에서 해결해야 할 문제를 다룬다. 해당 내용에는 마스터 데이터셋과 작업 파일을 관리하기 위하여 적절한 파일 구조 설정과 데이터 및 도큐멘테이션에 대한 안전한 백업을 제시하고 있다.

5단계는 '공유할 데이터 준비'로서 데이터 아카이빙 또는 배포를 준비하는 단계로서 아카이브된 기밀 데이터에 접근하고 분석하는 방법에 대한 정보를 제공한다. 세부적인 항목으로 공개위험 제한이 있는 데이터에 대한 문제를 해결하고 기탁할 파일 형식을 제안한다.

마지막 6단계는 '데이터 기탁'과 관련된 내용으로서 관련 양식을 작성하고 데이터의 배포 표준 및 형식을 준수하도록 안내하고 있다. 특히, 데이터셋의 아카이브를 위하여 다른 사람이 읽을 수 있는 방식으로 데이터셋을 구성해야 한다고 제시하고 있는데, 표준 통계 패키지를 사용하여 데이터셋에 접근할 수 있는 방법으로 (1) 표준 통계 프로그램으로 읽을 수 있도록 설정 파일과 함께 원시 ASCII 형식으로 데이터를 제공; (2) 특정 분석 프로그램 내에서 데이터를 시스템 파일로 제공; (3) 통계 프로그램에 의해 생성된 휴대용 파일로 데이터를 제공 등 세 가지 방법에 대해 안내하고 있다.

### 3.3 IWGDD(Interagency Working Group on Digital Data)

IWGDD(Interagency Working Group on Digital Data)는 디지털 과학 데이터에 대한 광범위한 관심을 반영하기 위하여 2006년 12월에 국립과학위원회(National Science and Technology Council of Science)에 의해 설립되었다. IWGDD의 목적은 연방정부가 과학, 기술, 공학 분야의 연구, 개발, 교육을 위한 디지털 데이터에 대한 신뢰성 있는 보존과 효과적인 접근을 보장하기 위해 상호운용 가능한 개방형 프레임워크에 관한 전략 계획의 추진을 개발하고 촉진하는 것이다(wikia.org 2019).

IWGDD의 보고서에 따르면, 전체 데이터 라이프 사이클에는 생성, 입수 또는 수집, 도큐멘테이션, 구성, 마이그레이션, 보호, 접근 그리고 이관 및 폐기(Disposition)을 포함하며 여기에는 두 가지 중요한 기능이 있다고 제시하고 있다. 다음은 두 가지 기능을 요약한 것이다(IWGDD 2009).

- 사이클은 정적인 것이 아니라 역동적이며 진행 중인 큐레이션, 이관 및 폐기(disposition) 및 이용 프로세스를 포함한다. 변환이나 재조합을 포함한 데이터 분석과 같은 많은 프로세스는 보존 및 접근을 위한 데이터의 양을 지속적으로 증가시킨다.
- 사이클의 단계는 독립적이지 않다. 즉, 데이터를 새로운 기술로 마이그레이션하지 않으면 나중에 해당 데이터에 접근할 수 없다. 데이터를 안정적으로 보존하고 접근하고 효율적으로 사용하려면 각 단계를 효과적으로 관리하고 라이프 사이클의 각 단

계를 조정해야 한다.

다음의 <표 5>는 IWGDD의 디지털 데이터 라이프 사이클에 대한 내용을 표로 나타낸 것이다.

라이프 사이클의 구성요소는 '계획', '생성', '보존', '기술 습득 및 구현', '이관 및 폐기(Disposition)', '문서화', '구성', '보호' 그리고 '접근' 등 9가지로 구성되어 있다. 이 중 '계획', '생성', '보존' 및 '이관 및 폐기'는 순차적이지만, '문서화', '구성', '보호' 및 '접근' 등의 기능은 데이터 라이프 사이클 모든 단계에서 발생한다.

먼저, '계획'은 생성하거나 수집해야 하는 데이터의 결정과 데이터 및 메타데이터의 형식과 품질에 대한 표준을 포함한다. '생성'은 라이프 사이클 동안 데이터를 관리하기 위한 작업 및 책임을 지정하는 것으로 데이터를 생성하거나 수집, 필요한 기간 동안 보관, 관리 및 접근 위치 등에 대한 내용을 포함하고 있다. 세 번째, '보존'은 데이터 업데이트와 그대로 유지하는 등의 데이터 구성 및 저장에 대한 내용을 담고 있고, '기술 습득 및 구현'은 노후화 극복 및 성능 향상을 위한 기술 업데이트, 스토리지 용량 확장 그리고 데이터 처리를 위한 새로운 기술 구현 등이 포함된다. 다음으로 '이관 및 폐기'는 데이터를 폐기할 것인지 또는 적합한 곳으로 전송할 것인지에 대한 내용이다. 여섯 번째 단계인 '문서화'는 데이터 콘텐츠, 형식, 메타데이터 등에 대한 표준 정의와 메타데이터 생성 및 유지 그리고 데이터 이력 문서화 등을 포함하고 있다. '구성'은 데이터 아키텍처 등의 구조 설계 및 구현, 표준 준수에 대한 내용을 가리키고 '보호'는 데이터의 품질 관리, 접근

제한, 신뢰성과 진본 보장 등에 대한 내용을 포함한다. 마지막 단계인 '접근'은 콘텐츠 및 품질 등에 대한 카탈로그 작성 및 설명, 이용 가능한 데이터에 대한 정보 배포와 검색, 분석, 배포, 표현 방법 등에 대한 지원으로 구성되어 있다.

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

DataONE은 기본 사이버 인프라 구축, 교육 및 봉사 활동 프로그램을 통해 다 분야 과학 데이터의 보존, 접근, 이용 및 재사용을 지원하기 위해 2009년에 미국 국립 과학 재단(National Science Foundation)의 편당을 통해 설립되었다. DataONE의 역할은 과학자가 생성한 생태 및 환경 데이터를 위한 과학 데이터 아카이빙을 제공하며, 해당 기관의 목표는 다중 규모, 다 분야 및 다국적 데이터를 보존하고 접근할 수 있도록 하는 것이다(Wikipedia 2019).

DataONE 워킹 그룹에서는 2013년에 '과학 연구에서 공공 참여를 위한 데이터 관리 가이드' 보고서를 발표하였는데 해당 내용은 데이터 라이프 사이클에 대한 소개, 데이터 라이프 사이클과 관련된 모범 사례 제공 그리고 데이터 관리 프로세스와 해당 도구에 대한 설명을 포함하고 있다(DataONE Public Participation in Scientific Research Working Group 2013). 다음의 <표 6>은 DataONE의 데이터 라이프 사이클 모델을 단계별로 정리한 것이다.

DataONE의 데이터 라이프 사이클은 '계획', '수집', '보장', '기술(Describe)', '보존', '탐색', '통합' 그리고 '분석' 등 8단계로 구성되어 있다.

<표 5> IWGDD의 디지털 데이터 라이프 사이클 모델(IWGDD 2009)

단계	내용	
계획	<ul style="list-style-type: none"> <li>연구 과제 또는 미션 기능을 지원하기 위해 생성하거나 수집해야 하는 데이터 결정</li> </ul>	<ul style="list-style-type: none"> <li>필요한 데이터의 기존 출처 식별 및 평가</li> </ul>
	<ul style="list-style-type: none"> <li>데이터와 메타데이터의 형식 및 품질에 대한 표준 식별</li> </ul>	
생성	<ul style="list-style-type: none"> <li>라이프 사이클 동안 데이터를 관리하기 위한 작업 및 책임 지정</li> <li>데이터를 생성 또는 수집</li> <li>필요한 기간 동안 관리 및 접근할 위치에 데이터를 저장</li> <li>파생 제품 생산: 예: 데이터 요약, 데이터 집계, 보고서, 출판물</li> </ul>	
보존	<ul style="list-style-type: none"> <li>데이터 구성 및 저장</li> </ul>	<ul style="list-style-type: none"> <li>기존 컬렉션에 업데이트 및 추가 사항 통합</li> <li>데이터를 필요한 기간 동안 그대로 유지</li> </ul>
기술 습득 및 구현	<ul style="list-style-type: none"> <li>노후화 극복 및 성능 향상을 위한 기술 업데이트</li> <li>필요에 따라 스토리지 및 처리 용량 확장</li> <li>데이터 수집, 처리, 분석, 검색 및 접근에 대한 진화하는 요구를 지원하는 새로운 기술 구현</li> </ul>	
이관 및 폐기	<ul style="list-style-type: none"> <li>중료 전략: 현재 리포지토리가 더 이상 유지할 수 없을 경우 다른 존재(entity)에 데이터 전송 계획</li> <li>일단 목적이 충족되면, 데이터를 폐기할 것인지 또는 다른 요구나 기회를 해결하기에 적합한 다른 조직으로 전송할 것인지 결정</li> </ul>	
문서화	<ul style="list-style-type: none"> <li>데이터 콘텐츠, 형식, 메타데이터, 품질, 업데이트 빈도 등에 대한 표준 정의</li> <li>메타데이터 생성 / 유지</li> <li>데이터 이력 문서화: 출처 및 이력, 실제 데이터 수집 및 처리 (예: 교정, 지리 참조, 소음 감소)</li> <li>이상 징후 및 누락(lacunae)에 주목</li> <li>이관, 폐기 결정 및 조치 기록</li> </ul>	
구성	<ul style="list-style-type: none"> <li>데이터 아키텍처, 엔지니어링 및 구조 설계 및 구현</li> <li>표준 준수</li> </ul>	
보호	<ul style="list-style-type: none"> <li>품질 관리 구현</li> </ul>	<ul style="list-style-type: none"> <li>수집 시 데이터 확인 및 검증</li> <li>변환 또는 파생 제품의 무결성과 유효성 보장</li> </ul>
	<ul style="list-style-type: none"> <li>접근 제한 구현</li> </ul>	<ul style="list-style-type: none"> <li>재산권 존중</li> <li>개인 정보 보호 및 기밀성 보호</li> </ul>
	<ul style="list-style-type: none"> <li>인증된 이용자에게 가용성 보장</li> </ul>	<ul style="list-style-type: none"> <li>이용자 역할 및 권한 정의</li> <li>개별 이용자 자격 부여</li> </ul>
	<ul style="list-style-type: none"> <li>신뢰성과 진본 보장</li> </ul>	<ul style="list-style-type: none"> <li>신뢰할 수 있는 리포지토리로서 기능</li> <li>시스템 및 시스템에 저장된 자산의 보안을 구현, 유지 및 모니터링</li> </ul>
<ul style="list-style-type: none"> <li>진위 확인 및 검증 방법 구현</li> </ul>		
접근	<ul style="list-style-type: none"> <li>기존 출처에서 데이터 수집</li> <li>콘텐츠, 품질, 가용성 등에 대한 카탈로그 작성 및 설명</li> <li>배포된 데이터에 대한 일관성 있는 식별 보장</li> <li>이용 가능한 데이터에 대한 정보 배포</li> <li>다양한 기술을 통해 다양한 이용 지원</li> <li>다양한 검색, 분석, 용도 변경, 배포, 표현 방법 지원</li> </ul>	

〈표 6〉 DataONE의 데이터 라이프 사이클 모델(DataONE Public Participation in Scientific Research Working Group 2013)

단계	내용	
계획(Plan)	<ul style="list-style-type: none"> <li>• 전체 데이터 라이프 사이클 동안 프로세스 및 정보자원을 계획</li> <li>• 프로젝트 목표(원하는 결과, 결과 및 영향)에서 시작하여 거꾸로 DMP(데이터 관리 계획), 지원 정책 및 지속 가능성 계획 수립</li> </ul>	<ul style="list-style-type: none"> <li>- DMP(데이터 관리 계획)</li> <li>- DMP Tool(데이터 관리 계획 도구)</li> </ul>
수집(Collect)	<ul style="list-style-type: none"> <li>• 참여자의 정보를 이용 가능한 데이터 파일로 가져 오는 가장 좋은 방법 결정</li> <li>• 이 결정 프로세스의 최종 결과는 데이터가 구성되는 방식을 설명하는 데이터 모델임</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 입력 및 처리(Manipulation)</li> </ul>
보장(Assure)	<ul style="list-style-type: none"> <li>• 데이터 품질을 향상시키는 품질 보증 및 품질 관리 절차(예: 교육 참가자, 일상적인 계측기 교정)를 사용하고 이를 해결하기 위한 잠재적 오류 및 기술 파악</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 품질 제어 및 보장</li> </ul>
기술(Describe)	<ul style="list-style-type: none"> <li>• 데이터의 이유, 사용자, 시기, 위치 및 방법을 설명하여 데이터 문서화</li> <li>• 메타데이터는 데이터 공유와 재사용의 핵심이며, 데이터를 기술하는 데 도움이 되는 표준이나 소프트웨어와 같은 많은 도구를 이용할 수 있음</li> </ul>	<ul style="list-style-type: none"> <li>- 메타데이터</li> <li>- 메타데이터 도구</li> </ul>
보존(Preserve)	<ul style="list-style-type: none"> <li>• (예를 들어 사고를 통한) 잠재적 손실을 최소화하기 위해 단기적으로 데이터를 보존하고, 장기적으로는 프로젝트 이해관계자 등이 향후 데이터에 접근, 해석 및 활용할 수 있도록 계획</li> <li>• 보존할 데이터, 보존할 위치 및 데이터와 함께 필요한 도큐멘테이션 결정</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 보호 및 백업</li> </ul>
탐색(Discover)	<ul style="list-style-type: none"> <li>• 프로젝트 데이터에 가치를 더할 수 있는 보완적인 데이터셋 식별</li> <li>• 데이터가 최대한 영향을 받도록 하는 전략에는 프로젝트 디렉토리 사이트에 프로젝트 등록, 오픈 리포지토리에 데이터 저장 및 메타데이터 정보센터에 데이터 설명 추가가 포함됨</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 공유</li> </ul>
통합(Integrate)	<ul style="list-style-type: none"> <li>• 시민 과학 프로젝트 데이터를 다른 데이터 소스와 결합하여 새로운 분석 및 조사가 가능함</li> <li>• 성공적인 데이터 통합은 데이터 라이프 사이클 내내 우수한 데이터 관리 방식을 채택하는 데 달려 있음</li> </ul>	<ul style="list-style-type: none"> <li>- 분석 및 워크플로우</li> </ul>
분석(Analysis)	<ul style="list-style-type: none"> <li>• 전문 연구자, 참여자 및 기타 이해 관계자의 프로젝트 목표를 충족하는 분석을 위해 시민 과학 프로젝트의 데이터 사용</li> <li>• 데이터 탐색, 분석 및 시각화를 지원하기 위해 많은 소프트웨어 도구를 사용할 수 있음</li> </ul>	<ul style="list-style-type: none"> <li>- 분석 체크리스트</li> <li>- 무료 분석 도구 예시</li> </ul>

첫 번째 단계인 '계획'은 DMP 등 전체 데이터 라이프 사이클 동안 프로세스 및 정보자원을 계획하는 내용을 두 번째 단계인 '수집'은 데이터

구성 방식을 설명할 수 있도록 참여자의 정보를 이용 가능한 데이터 파일 수집하기 위한 방법을 결정하는 것을 포함하고 있다. 세 번째는 '보

장' 단계로서 데이터 품질 제어 및 보장 등 데이터 품질을 향상시키는 품질 보증 및 품질 관리 절차를 제안하고 있으며, '기술(Describe)' 단계에서는 메타데이터 등과 같은 데이터의 이유, 사용자, 시기, 위치 및 방법을 설명하는 문서화에 대한 내용을 제시하고 있다. 다음으로 '보존'은 데이터 보호 및 백업을 위하여 데이터에 접근하고 해석할 수 있도록 단기 및 장기 보존 계획 수립을 '탐색(Discover)'은 데이터 공유를 위하여 해당 사이트에 프로젝트 등록, 오픈 리포지토리에 데이터 저장 등에 대한 내용을 포함하고 있다. 일곱 번째 단계인 '통합'은 프로젝트 데이터를 다른 데이터 소스와 결합하여 새로운 분석 및 조사가 가능하도록 분석 및 워크플로우에 대한 내용을 제시하고 있다. 마지막 단계인 '분석'은 데이터 분석을 위해 시민 과학 프로젝트의 데이터 사용과 분석 체크리스트와 무료 분석 도구를 제시함으로써 데이터 탐색, 분석 및 시각화를 지원하는 내용을 담고 있다.

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

USGS(U.S. Geological Survey)는 1879년에 설립된 미국 내무부의 유일한 과학 기관으로서 공공토지에 대한 지도를 만들고 지질 구조를 조사하며 광물 자원을 평가하는 것을 목표로 출발하였다. USGS의 과학자들은 미국의 환경, 천연 자원 및 자연 위험에 대해 연구하며, 주요 연구 분야는 생물학, 지리, 지질학 및 수문학 등이 있다. 2012년 이후로는, 다음과 같은 6가지 연구 분야에 대하여 연구하고 있다(USGS 2019).

- 기후 및 토지 이용 변화(Climatic and Land Use Change)
- 코어 사이언스 시스템(Core Science Systems)
- 생태계(Ecosystems)
- 에너지, 광물 및 환경(Energy and Minerals and Environmental Health)
- 자연 재해(Natural Hazards)
- 물(Water)

USGS에서는 데이터가 연구용도 이외에 잠재적 가치를 가진 기관 자산임을 나타내므로 라이프 사이클 내내 적절히 관리되어야 한다고 보고 있다. USGS 팀은 이러한 동기를 가지고 개념에서 보존 및 공유에 이르기까지 데이터에 대한 높은 수준의 관점으로 과학 데이터 라이프 사이클 모델(Science Data life cycle Model, SDLM)을 개발하였으며, 특히 USGS 과학자들은 해당 모델을 통해 데이터 제품이 잘 기술되고 보존되고 접근 가능하며 재사용에 적합하도록 보장할 수 있다고 하였다(USGS 2019). 다음의 <표 7>은 USGS의 데이터 라이프 사이클을 정리한 것으로 9가지 하위요소로 구성되어 있다.

USGS의 데이터 라이프 사이클은 순차적 단계인 '계획', '수집', '프로세스', '분석', '보존' 그리고 '출판/공유' 6개의 단계로 구성되어 있고 라이프 사이클 모든 단계에서 지속적으로 수행되어야 할 단계인 '기술(메타데이터, 문서화, 이션)', '품질 관리' 그리고 '백업 및 보안' 이렇게 3개의 단계를 포함하고 있다.

순차적 단계로 첫 번째 '계획'은 DMP 작성, 데이터 관리자의 역할 및 책임, 데이터 공유 계

〈표 7〉 USGS의 데이터 라이프 사이클(Faundeen, J. L. et al. 2013)

단계		세부 내용	
계획 (Plan)		<ul style="list-style-type: none"> <li>• 데이터 관리 계획을 작성하고 중요한 계획 활동에 대해 학습</li> <li>• 프로젝트를 시작하기 전에 라이프 사이클 동안 데이터를 관리하는 방법을 계획하는 것이 중요</li> </ul>	<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>
			<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>
수집 (Acquire)		<ul style="list-style-type: none"> <li>• 새로운 데이터 수집 또는 생성</li> <li>• 기존 데이터 확보</li> </ul>	<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>
			<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>
프로세스 (Process)		<ul style="list-style-type: none"> <li>• 데이터 처리에는 신규 또는 이전에 수집한 데이터 입력 준비와 관련된 다양한 활동이 포함됨</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 아카이빙</li> <li>- 데이터 이관 및 폐기(disposition)</li> <li>- 데이터 리포지토리</li> </ul>
			<ul style="list-style-type: none"> <li>- 데이터 공유 이유</li> <li>- USGS 데이터 릴리스 프로세스</li> <li>- 민감한 데이터</li> <li>- DOI(Digital Object Identifier)</li> <li>- 데이터 인용</li> <li>- 데이터 카탈로그</li> </ul>
분석 (Analyze)		<ul style="list-style-type: none"> <li>• 데이터 분석에는 처리된 데이터 탐색 및 해석과 관련된 다양한 활동이 포함됨</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 아카이빙</li> <li>- 데이터 이관 및 폐기(disposition)</li> <li>- 데이터 리포지토리</li> </ul>
			<ul style="list-style-type: none"> <li>- 데이터 공유 이유</li> <li>- USGS 데이터 릴리스 프로세스</li> <li>- 민감한 데이터</li> <li>- DOI(Digital Object Identifier)</li> <li>- 데이터 인용</li> <li>- 데이터 카탈로그</li> </ul>
보존 (Preserve)		<ul style="list-style-type: none"> <li>• 보존에는 데이터의 장기적인 생존 및 접근성을 보장하기 위해 사용되는 조치 및 절차가 포함</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 아카이빙</li> <li>- 데이터 이관 및 폐기(disposition)</li> <li>- 데이터 리포지토리</li> </ul>
			<ul style="list-style-type: none"> <li>- 데이터 공유 이유</li> <li>- USGS 데이터 릴리스 프로세스</li> <li>- 민감한 데이터</li> <li>- DOI(Digital Object Identifier)</li> <li>- 데이터 인용</li> <li>- 데이터 카탈로그</li> </ul>
출판/공유 (Publish/Share)		<ul style="list-style-type: none"> <li>• 데이터를 출판하고 공유하는 것은 동료들이 검토한 전통적인 저널 기사를 게재하는 것과 마찬가지로 연구 과정에서 중요하고 필요한 단계임</li> </ul>	<ul style="list-style-type: none"> <li>- 데이터 아카이빙</li> <li>- 데이터 이관 및 폐기(disposition)</li> <li>- 데이터 리포지토리</li> </ul>
			<ul style="list-style-type: none"> <li>- 데이터 공유 이유</li> <li>- USGS 데이터 릴리스 프로세스</li> <li>- 민감한 데이터</li> <li>- DOI(Digital Object Identifier)</li> <li>- 데이터 인용</li> <li>- 데이터 카탈로그</li> </ul>
Cross-Cutting 요소	기술 (메타데이터, 도큐멘테이션)	<ul style="list-style-type: none"> <li>• 데이터 라이프 사이클 동안 데이터에 대한 조치를 반영하도록 도큐멘테이션이 작성되어야 하며 업데이트해야 함</li> </ul>	<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>• 데이터 품질 관리는 과학 데이터 라이프 사이클의 모든 단계에서 데이터를 올바르게 수집, 처리, 처리, 사용 및 유지 관리하기 위해 프로토콜 및 방법을 사용하는 프로세스임</li> </ul>	<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 접근으로부터 데이터를 보호</li> </ul>	<ul style="list-style-type: none"> <li>- 손실, 손상 및 무단 접근으로부터 데이터를 보호</li> </ul>



약 등 프로젝트를 시작하기 전 라이프 사이클 동안 데이터를 관리하는 방법을 준비하는 내용을 포함한다. '수집'은 새로운 데이터의 수집 또는 생성 단계로서 데이터 수집을 위한 보안 요구 사항, 데이터 표준 준수, 파일 유형 선택 등 기존 데이터를 확보하는 방법에 대해 안내한다. 세 번째 단계인 '프로세스'는 실제 데이터를 처리하기 위한 내용으로 데이터의 검증, 요약, 변환, 통합, 파생 등으로 구성되어 있다. 다음 '분석'은 '프로세스' 단계에서 처리된 데이터의 해석에 관련한 활동으로 통계 분석, 시각화, 공간 분석, 모델링, 해석 등의 내용을 포함한다. 다음 '보존' 단계는 데이터 아카이빙, 데이터 이관 및 폐기(disposition), 데이터 리포지토리 등 데이터의 장기적인 생존 및 접근성을 보장하기 위해 사용되는 조치 및 절차로 구성된다. 여기에서 '데이터 이관 및 폐기'는 데이터가 더는 효용 가치가 없을 때 데이터의 폐기 또는 타 기관 및 리포지토리에 이관하는 것을 의미한다. 순차적 단계의 마지막인 '출판/공유'는 데이터를 공개하기 위한 절차, 민감 데이터의 처리, DOI 부여, 데이터 인용 등 데이터를 출판하고 공유하기 위한 내용을 나타낸다.

지속적 단계로서 '기술'은 데이터 라이프 사이클 동안 데이터에 대한 메타데이터와 도큐멘테이션의 작성 및 업데이트를 안내한다. 두 번째, '품질 관리'는 품질 보증 계획, 품질 관리 그리고 데이터 품질 문서화 등 데이터 라이프 사이클의 모든 단계에서 데이터를 올바르게 수집, 처리, 사용 및 유지 관리하기 위한 프로토콜 및 방법을 포함한다. 마지막으로 '백업 및 보안'은 우발적인 데이터 손실 및 무단 접근으로부터 데이터를 보호하기 위한 조치를 가리킨다.

### 3.6 UKDA(UK Data Archive)

영국 데이터 아카이브(UK Data Archive, 이하 UKDA)는 50년 역사의 전통을 가진 최대 규모의 사회 과학 및 인구 데이터의 수집, 큐레이팅 및 접근을 제공하는 센터로서 영국의 Essex 대학에서 운영되고 있다. UKDA의 기본적인 목적은 수집 및 생성된 데이터 및 메타데이터 관리에 대한 전문 지식을 통해 데이터를 오래 사용할 수 있고 이해할 수 있는 상태로 유지하는 것이며, 자세한 역할은 다음과 같이 요약될 수 있다(UK Data Archive 2019).

- 사회, 경제 및 인구 데이터 수집에 대한 지원, 교육 및 접근 제공
- 신뢰할 수 있는 디지털 리포지토리 제공
- 기술 데이터 인프라와 함께 데이터 거버넌스 및 정보 보안에 대한 전문 지식 제공
- 데이터 관리 분야에서 국제적인 모범 사례 공유
- 통계청, 사회 연구 센터 및 인구 조사 데이터에 의해 수행된 많은 조사를 포함하여 중요한 국가 데이터 자원 보유

다음의 <표 8>은 UKDA의 데이터 라이프 사이클 내용을 정리한 것이다.

UKDA가 제시한 데이터 라이프 사이클은 '데이터 생성', '데이터 처리', '데이터 분석', '데이터 보존', '데이터에 대한 접근 권한 부여', '데이터 재사용' 등 6개의 구성 요소를 포함하고 있다.

첫 번째, '데이터 생성'은 연구 설계, DMP 작성, 데이터 수집, 메타데이터 생성 등에 대한 세부 내용으로 구성되어 있다. 다음 '데이터 처리'

〈표 8〉 영국 데이터 아카이브의 데이터 라이프 사이클(VEERLE VAN DEN EYNDEN 2013)

단계	내용
데이터 생성	<ul style="list-style-type: none"> <li>• 연구 설계</li> <li>• 데이터 관리 계획(포맷, 스토리지 등)</li> <li>• 공유 동의 계획</li> <li>• 기존 데이터 탐색</li> <li>• 데이터 수집(실험, 관측, 측정, 시뮬레이션)</li> <li>• 메타데이터 캡처 및 생성</li> </ul>
데이터 처리	<ul style="list-style-type: none"> <li>• 데이터 입력, 디지털화, 복사, 번역</li> <li>• 데이터 확인, 검증, 정리</li> <li>• 데이터 익명화</li> <li>• 데이터 설명</li> <li>• 데이터 관리 및 저장</li> </ul>
데이터 분석	<ul style="list-style-type: none"> <li>• 데이터 해석</li> <li>• 데이터 도출(derive data)</li> <li>• 연구 결과물 생산</li> <li>• 저자 간행물</li> <li>• 보존을 위한 데이터 준비</li> </ul>
데이터 보존	<ul style="list-style-type: none"> <li>• 데이터를 최상의 형식으로 마이그레이션</li> <li>• 데이터를 적절한 매체로 마이그레이션</li> <li>• 데이터 백업과 저장</li> <li>• 메타데이터 및 도큐멘테이션 생성</li> <li>• 데이터 보존</li> </ul>
데이터에 대한 접근 권한 부여	<ul style="list-style-type: none"> <li>• 데이터 배포</li> <li>• 데이터 공유</li> <li>• 접근 제어</li> <li>• 저작권 설정</li> <li>• 데이터 홍보</li> </ul>
데이터 재사용	<ul style="list-style-type: none"> <li>• 후속 연구</li> <li>• 새로운 연구</li> <li>• 연구 검토 수행</li> <li>• 결과물에 대한 면밀한 조사</li> <li>• 교육 및 학습</li> </ul>

는 데이터 입력, 데이터 검증, 데이터 익명화, 데이터 관리 및 저장을 ‘데이터 분석’은 데이터 해석, 데이터 도출, 연구 결과 생산, 보존을 위한 데이터 준비 등의 내용을 나타낸다. 네 번째, 데이터를 최상의 형식으로 보존하기 위한 마이그레이션, 데이터 백업과 저장, 메타데이터 및 도큐멘테이션 생성 등 보존과 관련한 활동을 포함한다. 다섯 번째, ‘데이터에 대한 접근 권한 부

여’는 데이터 배포, 데이터 공유, 접근 제어, 저작권 설정 등 이용자가 실질적으로 데이터에 접근하기 위한 기능으로 구성되어 있다. 마지막으로 ‘데이터 재사용’은 후속 연구, 새로운 연구를 위한 데이터 재사용, 연구 검토 수행, 결과에 대한 면밀한 조사 등 데이터를 재사용하기 위한 목적 및 검토의 내용을 담고 있다.

### 3.7 데이터 라이프 사이클 비교 분석

본 절에서는 3장 1절부터 6절까지 기술한 DCC, ICPSR, IWGDD, DataONE, USGS, UK Data Archive 등의 데이터 라이프 사이클 구성 요소를 비교 분석하였다. 비교 분석은 DCC의 데이터 라이프 사이클을 기준으로 수행하였다. DCC를 선택한 이유는 DCC가 연구 데이터 관리를 위한 정책을 제공하는 기관으로서 분석된 기관 중 가장 먼저 데이터 라이프 사이클을 만들어 제시하였기 때문이다. 다음의 <표 9>는 DCC의 데이터 라이프 사이클을 기준으로 이미

분석된 5개 기관의 데이터 라이프 사이클 구성 요소를 비교 분석한 것이다. 분석 기준으로 제시된 DCC의 데이터 라이프 사이클은 본 장의 1절에서 기술된 것처럼 3개의 활동 15개의 하위 요소로 구성되어 있다. DCC의 라이프 사이클 활동 중에서 '기술 및 표현정보'는 메타데이터에 대한 기술로서 순차적 활동의 '생성 및 입수'의 세부 내용과 일치하여 기준에서 삭제하였다. 또한 '전체 라이프 사이클 활동'의 '커뮤니티 활동 및 참여'는 관련된 커뮤니티가 공유 표준, 도구 및 적합한 소프트웨어 개발에 참여하는 것을 나타내므로 분석 기준에서 제외하였

<표 9> 데이터 라이프 사이클 비교 분석

기관 DCC 구성 요소	ICPSR	IWGDD	DataONE	USGS	UKDA	비고
개념화	• 제안서 개발 및 DMP • 프로젝트 시작	• 계획	• 계획	• 계획	• 데이터 생성	◎
생성 및 입수	• 데이터 수집 및 파일 생성	• 생성 • 문서화 • 접근	• 수집 • 기술(Describe)	• 수집 • 기술(Describe)	• 데이터 생성	◎
평가 및 선택			• 보존			
수집	• 데이터 기탁					
보존 활동	• 데이터 기탁 • 공유할 데이터 준비	• 보존 • 보호	• 보장 • 보존	• 보존 • 품질관리 • 백업 및 보안	• 데이터 처리 • 데이터 보존	◎
저장	• 데이터 분석	• 보존	• 보존	• 보존	• 데이터 보존	◎
접근, 이용 및 재사용		• 접근 • 기술 습득 및 구현	• 탐색	• 출판/공유	• 데이터에 대한 접근 권한 부여 • 데이터 재사용	◎
변환		• 생성 • 보호	• 보존	• 프로세스 • 보존	• 데이터 보존	◎
폐기		• 기관 및 폐기		• 보존		◎
재평가			• 통합			
			• 분석	• 분석	• 데이터 처리	◎

고, '보존 계획'도 순차적 활동에 '보존 활동'이 있어 해당 기준에서 제거하기로 하였다. 여기에 '큐레이션 및 보존'도 전체 라이프 사이클 활동의 관리 및 운영을 나타내는바 순차적 활동의 '개념화'의 내용과 일치한다고 판단하여 제외하기로 하였다. 마지막으로 '비정기적 활동'의 마이그레이션은 순차적 활동의 '변환' 단계와 내용이 유사하여 기준에서 삭제하였다. 최종적으로 각 데이터 라이프 사이클을 분석하기 위한 기준 요소는 '개념화', '생성 및 입수', '평가 및 선택', '수집', '보존 활동', '저장', '접근, 이용 및 재사용', '변환', '폐기, 그리고 '재평가' 등 10개로 구성하였다.

비교 분석한 결과 DCC를 포함하여 6개 모든 기관에 포함된 구성 요소는 '개념화', '생성 및 입수', '보존 활동' 그리고 '저장' 등 4개이며, 5개 기관에 포함된 구성 요소는 '접근, 이용 및 재사용' 그리고 '변환' 등 2개이다. 다음 DCC를 포함한 3개 기관에 포함된 구성 요소는 '폐기'이며, 2개의 기관에 포함된 구성 요소는 '평가 및 선택' 그리고 '수집'이다. DCC를 제외하고 어떠한 기관에서도 포함하지 않은 구성 요소는 '재평가' 1개의 항목이 도출되었다. 분석의 기준 요소로 사용되지 않았지만, DataONE의 '분석'은 USGS에서 '분석'이라는 동일 요소로 UKDA에서도 '데이터 처리'라는 유사한 요소로 3개 이상의 기관에 포함하고 있는 것을 확인하였다. 위의 <표 9>에서는 기준으로 제시된 DCC를 포함하여 3개 기관에 포함된 구성 요소에 '◎' 표시를 하여 구분하였고, 본 연구에서는 데이터 라이프 사이클을 제안하기 위한 최종 구성 요소로 DCC를 포함한 3개 기관 이상에 포함된 구성 요소를 제시하였다. 제시된 구성 요소는

'개념화', '생성 및 입수', '분석', '보존 활동', '저장', '접근, 이용 및 재사용', '변환' 그리고 '폐기' 등 8개이다.

이 중 '개념화'는 DCC를 제외한 다른 기관에서 '계획'이라는 구성 요소 명을 사용하고 있어 본 연구에서도 '개념화' 대신 '계획'이라는 용어로 제시하였다. '생성 및 입수'의 경우 ICPSR, DataONE 그리고 USGS에서는 '입수'보다 '수집'이라는 용어를 사용하여 해당 항목을 '생성 및 수집'으로 변경하였다. DCC의 구성요소에는 포함하고 있지 않은 '분석'은 DataONE에서 데이터 탐색, 분석 및 시각화를 지원하는 소프트웨어를 안내하며 USGS에서는 데이터 탐색 및 해석과 관련된 활동 그리고 UKDA에서는 '데이터 처리'라는 구성요소를 통해 데이터 해석 등에 대한 내용을 포함하고 있다. 여기에 USGS에서는 데이터를 분석하는 '분석' 이외에 '프로세스'라는 데이터 처리 항목을 포함하고 있어 '분석' 대신에 '프로세스'라는 구성요소로 대신하고자 한다. '보존 활동'과 '저장'은 DCC에서는 구분하여 2개의 구성 요소로 제시하지만, IWGDD, DataONE, USGS 그리고 UKDA에서는 보존의 의미에 저장의 의미를 포함하고 있어 해당 항목을 '보존'이라는 용어로 통일하고자 한다. 더불어 '보존 활동'은 보존의 의미 이외에 무결성 등의 보호 및 품질관리, 보존을 위한 조치로써 백업 및 보안, 진본과 신뢰성을 유지하기 위한 보장 등의 의미를 내포하고 있다. 이에 본 연구에서는 '보존활동'의 구성 요소를 '보존', '보장' 그리고 '백업 및 보안'으로 나누어 해당 세부내용을 제시하고자 한다. '접근, 이용 및 재사용'은 의미의 명확성을 위하여 '접근 및 이용'이라는 항목으로 정리하였다. 또한 '기

술(Description)'처럼 메타데이터에 대한 항목에 대해서 DCC와 USGS에서는 모든 라이프 사이클 단계에서 적용될 수 있는 세부 내용으로 제시하고 있고, ICPSR에서 '데이터 수집 및 파일 생성' 구성요소에 IWGDD에서는 '계획'과 '문서화' 구성요소에서 DataONE은 '기술'과 '탐색' 구성요소에 포함되어 있다. 이처럼 메타데이터는 2개의 구성요소에 포함된 것을 확인할 수 있으므로 '기술' 단계를 또 하나의 구성 요소로 제시하였다. 다음 DCC의 기준 중 하나인 '변환'은 다른 매체로의 마이그레이션을 의미하고 있어 ICPSR 이외의 기관에서 '생성', '보호', '보존' 그리고 '프로세스' 다양한 구성요소의 세부내용에 포함되어 있다. 각 기관에서 의미하는 '변환'의 의미를 본 연구에서는 USGS의 '프로세스'에서 일어나는 변환의 의미와 가장 유사하다고 판단하여 '프로세스'의 구성요소로 대신 제안하고자 한다. 그다음 DCC의 '폐기'는 선택되지 않은 데이터의 이관 및 활용가치가 없는 데이터 폐기에 대한 내용으로 IWGDD에서는 '이관 및 폐기(disposition)' 그리고 USGS에서는 '보존'의 세부내용에 이관 및 폐기의 내용을 포함하고 있다. 이에 따라 본 연구에서는 DCC의 '폐기'의 의미가 단순히 데이터 폐기만을 의미할 수 있어 IWGDD가 제시한 '이관 및 폐기(Disposition)'로 제시하고자 한다.

요약하여 본 연구에서 제시하고자 하는 데이터 라이프 사이클의 구성요소를 '계획', '생성 및 수집', '프로세스', '보존', '이관 및 폐기', '접근 및 이용', '기술', '보장' 그리고 '백업 및 보안' 등 9개 항목으로 제시하고자 한다. 다음의 4장에서는 9개의 구성요소를 기준으로 DCC, ICPSR, IWGDD, DataONE, USGS 그리고 UKDA에

서 제시된 내용을 분석하고자 한다.

## 4. 구성 요소별 세부 내용 비교

이번 장에서는 3장에서 도출한 데이터 라이프 사이클의 구성 요소인 '계획', '생성 및 수집', '프로세스', '보존', '이관 및 폐기', '접근 및 이용', '기술', '보장' 그리고 '백업 및 보안' 등 9가지 구성요소에 대하여 해당 내용을 기관별로 비교하여 제시하였다.

### 4.1 계획(Plan)

데이터 라이프 사이클의 구성요소인 계획(Plan)은 제안서 및 데이터 관리 계획(Data Management Plan) 작성을 나타낸다. 데이터 관리 계획(Data Management Plan)은 연구 중에 수집되거나 생성될 데이터를 설명하며, 여기에는 데이터 관리, 설명 및 저장 방법, 사용할 표준, 프로젝트 완료 중 그리고 완료 후 데이터 처리와 보호 방법을 포함한다(USGS 2019).

DCC 등 6개 기관 모두 '계획' 단계에서 DMP 작성에 대한 내용을 포함하고 있으며 ICPSR은 DMP와 더불어 연구의 제안서 개발에 대한 내용을 담고 있다. IWGDD는 메타데이터의 형식에 대한 내용을 DataONE은 라이프 사이클 동안 정보자원의 계획을 USGS는 데이터 관리자의 역할 및 책임, 데이터 공유 계약 개발 등을 포함하고 있다. 마지막으로 UKDA는 DMP 이외에 연구 설계와 공유 동의 계획 등을 제시하고 있다.

〈표 10〉 ‘계획’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
계획	<ul style="list-style-type: none"> <li>• 데이터 생성을 고안하고 계획</li> </ul>	<ul style="list-style-type: none"> <li>• 제안서 개발</li> <li>• DMP 작성</li> </ul>	<ul style="list-style-type: none"> <li>• 생성하거나 수집해야 하는 데이터 결정</li> <li>• 데이터, 메타데이터의 형식 및 품질에 대한 표준 식별</li> </ul>	<ul style="list-style-type: none"> <li>• 라이프 사이클 동안 프로세스 및 정보자원을 계획</li> <li>• DMP, 지원 데이터 정책 및 지속가능성 계획 수립</li> </ul>	<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>	<ul style="list-style-type: none"> <li>• 연구 설계</li> <li>• DMP 작성</li> <li>• 공유 동의 계획</li> </ul>

4.2 생성 및 수집(Create & Collect)

‘생성 및 수집’은 데이터 등 데이터셋을 직접 생성하거나 기존 데이터를 수집하는 단계로서 데이터 수집 방법까지 포함할 수 있다. 다음의 〈표 11〉은 기관별 ‘생성 및 수집’에 대한 세부 내용을 비교한 것이다.

6개 기관에서 제시된 ‘생성 및 수집’의 세부 내용을 요약하면 데이터 생성 및 수집 그리고 각 데이터에 대한 메타데이터 할당으로 정리될

수 있다.

4.3 프로세스(Process)

세 번째 구성요소인 ‘프로세스(Process)’는 데이터의 처리 및 분석에 관한 내용으로서 각 기관별 세부내용인 다음의 〈표 12〉와 같다.

DCC, ICPSR 그리고 IWGDD에서는 데이터의 처리 및 분석에 관련한 내용을 제시하지 않고 DataONE, USGS, UKDA에서만 포함하고

〈표 11〉 ‘생성 및 수집’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
생성 및 수집	<ul style="list-style-type: none"> <li>• 데이터 작성자, 기타 아카이브, 리포지토리 또는 데이터 센터에 의해 데이터를 수집</li> <li>• 필요한 경우 메타데이터 할당</li> </ul>	<ul style="list-style-type: none"> <li>• 정량적 데이터 및 정성적 데이터 수집</li> <li>• 해당 데이터의 메타데이터 및 도큐멘테이션 기술</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 생성 또는 수집</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터를 수집하기 위한 가장 좋은 방법 결정</li> </ul>	<ul style="list-style-type: none"> <li>• 새로운 데이터 수집 또는 생성</li> <li>• 기존 데이터 확보</li> </ul>	<ul style="list-style-type: none"> <li>• 기존 데이터 탐색</li> <li>• 데이터 수집(실현, 관측, 측정, 시물레이션)</li> <li>• 메타데이터 캡처 및 생성</li> </ul>

〈표 12〉 ‘프로세스’ 구성 요소의 세부 내용 비교

구성 요소	DataONE	USGS	UKDA
프로세스	<ul style="list-style-type: none"> <li>• 데이터 탐색, 분석 및 시각화를 지원하는 소프트웨어 도구 제시</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 처리(검증, 요약, 변환, 통합, 파생 등)</li> <li>• 데이터 분석(통계 분석, 시각화, 공간 및 이미지 분석, 모델링, 해석 등)</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 처리(데이터 입력, 검증, 익명화, 설명 등)</li> <li>• 데이터 분석(해석, 도출, 생산, 보존 준비 등)</li> </ul>

있는 것을 볼 수 있다. 이러한 ‘프로세스’는 생성 또는 수집된 데이터를 처리 그리고 분석하는 단계로서 데이터 처리에는 실제 데이터를 입력 준비와 관련된 데이터 검증, 요약, 변환, 익명화, 통합 등의 다양한 활동을 포함한다. 데이터 분석에는 분석 소프트웨어 제공부터 통계 분석, 시각화, 공간 및 이미지 분석, 모델링, 해석 등을 제시한다.

#### 4.4 보존(Preserve)

네 번째 구성요소인 ‘보존’은 데이터를 장기 간 보존하기 위한 조치와 접근성을 보장하기 위해 사용되는 절차 및 계획을 포함한다. 다음의 <표 13>은 ‘보존’에 대한 세부내용을 기관별로 정리한 것이다.

분석된 6개 기관에서 제시된 ‘보존’의 세부

내용은 데이터의 장단기 보존을 위한 계획 수립, 보존할 데이터 및 도큐멘테이션 결정, 보존 메타데이터의 생성 및 할당, 데이터의 장기 접근을 보장하기 위한 데이터 아카이빙, 데이터 리포지토리 안내 등을 포함한다.

#### 4.5 이관 및 폐기(Dispose)

기관 내에서 효용성을 다하거나 더 이상 필요가 없는 데이터의 이관 그리고 보존을 위한 데이터 선택 시 포함되지 않은 데이터의 폐기를 의미하는 ‘이관 및 폐기(Dispose)’는 본 연구에서 제시하는 데이터 라이프 사이클의 다섯 번째 구성요소이다. 다음의 <표 14>는 ‘이관 및 폐기’ 구성요소의 기관별 세부 내용을 비교한 것이다.

‘이관 및 폐기’는 DCC, IWGDD 그리고 USGS

<표 13> ‘보존’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
보존	<ul style="list-style-type: none"> <li>• 데이터의 관리 특성을 장기간 보존하고 보유하기 위한 조치</li> <li>• 보존 메타데이터 할당, 표현 정보 할당</li> </ul>	<ul style="list-style-type: none"> <li>• 파일 아카이브</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 구성 및 저장</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터를 단기적으로 보존할 계획 수립</li> <li>• 보존할 데이터, 보존할 위치 및 데이터와 함께 필요한 도큐멘테이션 결정</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터의 장기적인 생존 및 접근성을 보장하기 위해 사용되는 조치 및 절차(데이터 아카이빙, 데이터 리포지토리 안내)</li> </ul>	<ul style="list-style-type: none"> <li>• 메타데이터 및 도큐멘테이션 생성</li> <li>• 데이터 보존</li> </ul>

<표 14> ‘이관 및 폐기’ 구성 요소의 세부 내용 비교

구성 요소	DCC	IWGDD	USGS
이관 및 폐기	<ul style="list-style-type: none"> <li>• 큐레이션 및 보존을 위해 선택되지 않은 데이터 폐기</li> <li>• 다른 아카이브, 리포지토리, 데이터 센터 또는 기타 관리자에게 이관될 수 있음</li> </ul>	<ul style="list-style-type: none"> <li>• 리포지토리가 더 이상 유지할 수 없을 경우 다른 존재(entity)에 데이터 전송 계획</li> <li>• 데이터를 폐기할 것인지 또는 다른 요구나 기회를 해결하기에 적합한 다른 조직으로 전송할 것인지 결정</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 이관 및 폐기</li> </ul>

에서만 세부내용을 포함하고 있고, ICPSR, Data ONE 그리고 USGS에서는 제시하지 않고 있다. 해당 내용은 기관 내에서 큐레이션 및 보존을 위해 선택되지 않은 데이터를 폐기하거나 타 아카이브 등으로의 이관을 포함한다.

#### 4.6 접근 및 이용(Access & Use)

‘접근 및 이용(Access & Use)’은 데이터 접근, 이용 및 재사용과 관련한 내용으로 데이터 탐색, 공유 그리고 출판 등에 대한 내용으로 구성되어 있다. 다음의 <표 15>는 기관별 세부 내용을 비교한 것이다.

DCC는 접근, 이용 및 재사용 단계에서 이용자가 데이터에 접근할 수 있는지 확인을 ICPSR은 기탁 단계에서 허용되는 파일 형식 제시 그리고 IWGDD는 데이터 카탈로그 작성, 배포된 데이터의 식별 보장, 데이터에 대한 정보 배포 등 데이터에 접근하기 위한 다양한 내용을 포함하고 있다. 또한 DataONE은 ‘탐색’ 단계를 통해 프로젝트 사이트에 해당 프로젝트 등록, 오픈 리포지토리에 데이터 저장 등을 USGS는 출판 및 공유

단계에서 데이터 공개 프로세스, DOI, 데이터 인용 및 카탈로그에 대한 내용을 마지막으로 UKDA는 접근 권한 부여 및 재사용 단계에서 데이터 배포 공유, 접근 제어, 저작권 설정, 데이터 홍보 등에 대한 내용을 제시하고 있다.

#### 4.7 기술(Describe)

‘기술(Describe)’은 메타데이터의 생성과 도큐멘테이션 작성에 관한 내용으로 다음의 <표 16>과 같이 모든 기관에서 해당 내용을 다루고 있다.

먼저, DCC는 생성 및 수집, 보존 단계에서 메타데이터를 할당하도록 제시하고 있고, ICPSR은 프로젝트 시작, 수집 및 생성 단계에서 메타데이터 및 도큐멘테이션 기술을 IWGDD는 계획과 문서화 단계에서 메타데이터에 대한 표준 정의 그리고 메타데이터 생성 및 유지에 대한 내용을 포함하고 있다. 또한, DataONE은 기술 및 보존 단계에서 데이터 문서화, 메타데이터 및 메타데이터 작성 도구에 대한 안내를 포함하며, USGS는 모든 라이프 사이클 동안 메타

<표 15> ‘접근 및 이용’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
접근 및 이용	<ul style="list-style-type: none"> <li>지정된 이용자와 재사용자 모두가 매일 데이터에 액세스 할 수 있는지 확인</li> <li>접근 제어 및 인증 절차가 적용</li> </ul>	<ul style="list-style-type: none"> <li>허용되는 파일 형식 제시</li> </ul>	<ul style="list-style-type: none"> <li>카탈로그 작성 및 설명</li> <li>배포된 데이터에 대한 일관성 있는 식별 보장</li> <li>데이터에 대한 정보 배포</li> <li>다양한 이용 지원</li> <li>검색, 분석, 용도 변경, 배포, 표현 방법 지원</li> </ul>	<ul style="list-style-type: none"> <li>프로젝트 사이트에 프로젝트 등록</li> <li>오픈 리포지토리에 데이터 저장 및 메타데이터 정보센터에 데이터 설명</li> </ul>	<ul style="list-style-type: none"> <li>데이터 공개 프로세스</li> <li>DOI(Digital Object Identifier)</li> <li>데이터 인용</li> <li>데이터 카탈로그</li> </ul>	<ul style="list-style-type: none"> <li>데이터 배포</li> <li>데이터 공유</li> <li>접근 제어</li> <li>저작권 설정</li> <li>데이터 홍보</li> </ul>



〈표 16〉 ‘기술’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
기술	<ul style="list-style-type: none"> <li>• 관리, 설명, 기술, 구조 및 보존 메타데이터를 할당</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터셋 및 프로젝트 도큐멘테이션 기술</li> <li>• XML, DDI 등 메타데이터 생성 모범 사례</li> </ul>	<ul style="list-style-type: none"> <li>• 메타데이터에 대한 표준 정의</li> <li>• 메타데이터 생성 및 유지</li> <li>• 데이터 이력 문서화</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터의 이유, 사용자, 시기, 위치 및 방법을 설명하여 데이터 문서화</li> <li>• 메타데이터 및 메타데이터 작성 도구</li> <li>• 보존 시 필요한 도큐멘테이션 결정</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 라이프사이클 동안 메타데이터 및 도큐멘테이션 생성</li> </ul>	<ul style="list-style-type: none"> <li>• 메타데이터 캡처 및 생성</li> <li>• 메타데이터 및 도큐멘테이션 생성</li> </ul>

데이터 및 도큐멘테이션이 생성되어야 한다고 마지막으로 UKDA는 생성 및 보존 단계에서 메타데이터와 도큐멘테이션 생성을 제시하고 있다. 특히, 3장에서 기술한 바와 같이 ‘기술’은 라이프 사이클의 한 단계에서 실행되어야 할 내용보다는 모든 단계에서 제시해야 할 내용으로 표현하고 있다.

신뢰성에 대한 내용을 ICPSR은 데이터 생성 및 수집 단계에서 데이터셋에 대한 무결성을 IWGDD는 보호 단계에서 품질 관리와 신뢰성 및 진본 보장을 DataONE에서는 품질 관리 단계에서 품질 보증 계획, 품질 관리 및 데이터 품질 문서화 등을 포함하고 있다. 특히, USGS에서는 해당 단계를 모든 라이프 사이클 동안 이루어져야 할 활동으로 제시하고 있다.

#### 4.8 보장(Assure)

‘보장(Assure)’은 데이터의 무결성, 진본 및 신뢰성 등에 대한 내용으로 다음의 〈표 17〉을 통하여 기관별 세부내용을 비교하였다.

UKDA를 제외한 5개 기관에서 제시하고 있는 ‘보장’은 데이터의 품질관리에 대한 내용으로 DCC에서는 보존단계에서 무결성, 진본 및

#### 4.9 백업 및 보안(Backup & Secure)

본 연구에서 제안하는 마지막 구성요소인 ‘백업 및 보안(Backup & Secure)’은 실질적인 데이터 보호와 관련한 내용이다. 다음의 〈표 18〉은 기관별 해당하는 ‘백업 및 보안’의 내용을 비교한 것이다.

〈표 17〉 ‘보장’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS
보장	<ul style="list-style-type: none"> <li>• 데이터의 무결성을 유지하면서 진본, 신뢰성 및 가용성을 보장</li> <li>• 유효성 검사, 허용 가능한 데이터 구조 또는 파일 형식 보장</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터셋 무결성</li> </ul>	<ul style="list-style-type: none"> <li>• 품질 관리</li> <li>• 신뢰성과 진본 보장</li> <li>• 진위 확인 및 검증 방법</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 품질을 향상시키는 품질 보증 및 품질 관리 절차</li> </ul>	<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>

〈표 18〉 ‘백업 및 보안’ 구성 요소의 세부 내용 비교

구성 요소	DCC	ICPSR	IWGDD	DataONE	USGS	UKDA
백업 및 보안	<ul style="list-style-type: none"> <li>• 관련 표준을 준수하는 안전한 방식으로 데이터 저장</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 백업</li> <li>• 공개 위험 제한 원칙</li> <li>• 기밀 보호 관행</li> <li>• 이용이 제한된 데이터 수집</li> <li>• 데이터 Enclave</li> </ul>	<ul style="list-style-type: none"> <li>• 접근 제한을 통한 개인 정보 보호 및 기밀성 보호</li> </ul>	<ul style="list-style-type: none"> <li>• 잠재적인 손실 최소화를 위한 단기전 보존 계획 수립</li> </ul>	<ul style="list-style-type: none"> <li>• 우발적인 데이터 손실, 손상 및 무단 접근으로부터 데이터를 보호</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 익명화</li> <li>• 데이터 백업과 저장</li> </ul>

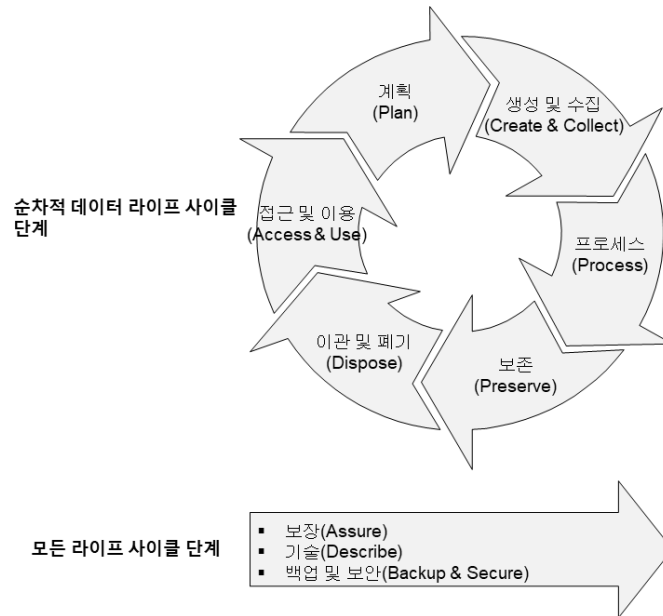
DCC는 저장 단계에서 안전한 방식으로 데이터의 저장을 ICPSR은 분석과 준비 단계에서 데이터 백업, 공개 위험 제한, 기밀 보호, 이용 제한 데이터 수집 등을 IWGDD는 보호 단계에서 개인 정보 보호 및 기밀성 보호에 대한 내용을 제시하고 있다. 또한, DataONE은 보존 단계에서 잠재적인 손실을 최소화하기 위한 단기 보존 계획 수립을 USGS는 백업 및 보안 단계에서 데이터 손실, 손상 및 무단 접근으로부터 데이터의 보안을 마지막 UKDA에서는 처리와 보존 단계에서 데이터 익명화, 데이터 백업 및 저장에 대한 내용을 포함하고 있다. 특히, ‘백업 및 보안’은 USGS에서 모든 라이프 사이클 동안 수행되어야 할 내용으로 제시하고 있으며 다른 기관에서도 저장, 보존, 처리 등 다양한 단계의 세부 내용으로 제시하고 있는 것을 확인할 수 있다.

### 5. 연구 데이터 관리를 위한 데이터 라이프 사이클 모델 제안

이번 장에서는 4장에서 논의한 각 기관의 데이터 라이프 사이클의 세부 내용을 중심으로 연구 데이터 관리 및 서비스를 지원하기 위한

데이터 라이프 사이클을 제안하였다. 즉, 6개의 기관에서 제시한 데이터 라이프 사이클의 구성 요소와 해당 요소의 세부 내용을 중심으로 문헌 분석을 통해 모델을 제시하였다. 기존 모델에서는 해당 분야 등을 중심으로 데이터 라이프 사이클을 제시하였다면, 본 연구에서 도출된 모델은 모든 분야에서 연구 데이터 관리를 위한 데이터 라이프 사이클로 제시될 수 있다.

다음의 〈그림 2〉는 제안된 데이터 라이프 사이클 모델을 도식화한 것이다. 본 연구에서 제안한 데이터 라이프 사이클은 2개의 영역으로 구분하였다. 첫 번째는 ‘순차적 데이터 라이프 사이클 단계’를 제시한 것으로 순서대로 일어나는 라이프 사이클을 도식화한 것이다. 순차적 라이프 사이클 단계는 ‘계획(Plan)’, ‘생성 및 수집(Create & Collect)’, ‘프로세스(Process)’, ‘보존(Preserve)’, ‘이관 및 폐기(Dispose)’ 그리고 ‘접근 및 이용(Access & Use)’ 등 6개의 단계로 구성되어 있다. 두 번째는 ‘모든 라이프 사이클 단계’로서 ‘보장(Assure)’, ‘기술(Describe)’ 그리고 백업 및 보안(Backup & Secure) 등 3가지의 구성 요소를 포함하고 있다. ‘모든 라이프 사이클 단계’는 라이프 사이클 모든 단계에서 수행되어야 할 조치, 행동 등에 관한 내용을 제시하였다. 다음의 〈그림 2〉는 본 연구에서



〈그림 2〉 연구 데이터 관리를 위한 데이터 라이프 사이클

제시한 연구 데이터 관리를 위한 데이터 라이프 사이클을 도식화한 것이다.

다음의 〈표 19〉는 연구 데이터 관리를 위한 데이터 라이프 사이클에 포함되어야 할 구성 요소 및 세부 내용을 기술한 것이다.

본 연구에서 제시하고자 하는 연구 데이터 관리 및 서비스를 위한 데이터 라이프 사이클은 2개의 영역(순차적 단계, 모든 라이프 사이클 단계)으로 구성하였다. 첫 번째, 순차적 단계는 계획, 생성 및 수집, 프로세스, 보존, 이관 및 폐기, 접근 및 이용 등 6개의 단계로 두 번째, 모든 라이프 사이클 단계는 기술, 보장, 백업 및 보안 등 3개의 단계로 구성하였다. 단계별 세부 구성 요소는 ‘계획’에서 제안서 및 DMP 작성, ‘생성 및 수집’에서는 생성과 수집으로 ‘프로세스’는 처리와 분석, ‘접근 및 이용’은 접근 및 이용과 출판으로 구성하였고 ‘기술’은 메타데이터

와 문서화(Documentation)로 ‘백업 및 보안’은 백업과 보호로 제시하였지만 ‘보존’, ‘이관 및 폐기’ 그리고 ‘보장’은 해당 내용을 따로 세분화하지 않았다. 해당 구성 요소의 세부 내용은 〈표 19〉에 상세히 기술되어 있다.

## 6. 결론

기술 및 연구 측정 도구의 발전, 데이터 저장 용량 증가, 연구 데이터의 접근 및 검색 가능성 향상으로 데이터 공유, 통합 및 재사용에 크게 의존하는 새로운 연구 환경이 조성되었다. 이러한 환경에서 성공하려면 데이터 사이클 동안 연구 데이터를 의도적으로 관리해야 한다(Carnegie Mellon University Libraries 2019). 미국, 영국 그리고 호주 등 선진국을 중심으로 연구 데

〈표 19〉 데이터 라이프 사이클의 단계별 세부 내용

구분	단계	구성 요소	세부 내용
순차적 단계	계획 (Plan)	제안서 및 DMP 작성	<ul style="list-style-type: none"> <li>• 연구 설계</li> <li>• 데이터 생성을 고안하고 계획</li> <li>• 생성하거나 수집해야 하는 데이터 결정</li> <li>• 데이터, 메타데이터의 형식 및 품질에 대한 표준 식별</li> <li>• 라이프 사이클 동안 프로세스 및 정보자원을 계획</li> <li>• 데이터 관리자의 역할 및 책임</li> <li>• 데이터 공유 계약 개발</li> <li>• 데이터 접근 제어 및 저작권 이해</li> <li>• 공유 동의 계획</li> </ul>
	생성 및 수집 (Create & Collect)	생성	<ul style="list-style-type: none"> <li>• 데이터를 생성하기 위한 가장 좋은 방법 결정</li> <li>• 데이터 생성</li> <li>• 메타데이터 할당 및 생성</li> </ul>
		수집	<ul style="list-style-type: none"> <li>• 기존 데이터 탐색 및 확보</li> <li>• 데이터 수집(실험, 관측, 측정, 시뮬레이션)</li> <li>• 정량적 데이터 및 정성적 데이터 수집</li> <li>• 데이터 작성자, 기타 아카이브, 리포지토리 또는 데이터 센터에 의해 데이터를 수집</li> <li>• 해당 데이터의 메타데이터 및 도큐멘테이션 기술</li> </ul>
	프로세스 (Process)	처리	<ul style="list-style-type: none"> <li>• 데이터 입력</li> <li>• 데이터 검증</li> <li>• 데이터 요약</li> <li>• 데이터 변환</li> <li>• 데이터 통합</li> <li>• 데이터 파생</li> <li>• 데이터 익명화</li> <li>• 데이터 설명</li> <li>• 데이터 처리를 위한 소프트웨어 제시</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> <li>• 데이터 분석을 위한 소프트웨어 제시</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>
		이관 및 폐기 (Dispose)	<ul style="list-style-type: none"> <li>• 데이터를 폐기할 것인지 또는 다른 요구나 기회를 해결하기에 적합한 다른 조직으로 전송할 것인지 결정</li> <li>• 큐레이션 및 보존을 위해 선택되지 않은 데이터 폐기</li> <li>• 다른 아카이브, 리포지토리, 데이터 센터 또는 기타 관리자에게 이관</li> </ul>

구분	단계	구성 요소	세부 내용
순차적 단계	접근 및 이용 (Access & Use)	접근 및 이용	<ul style="list-style-type: none"> <li>• 지정된 이용자와 재사용자 모두가 매일 데이터에 액세스 할 수 있는지 확인</li> <li>• 접근 제어 및 인증 절차가 적용</li> <li>• 허용되는 파일 형식 제시</li> <li>• 프로젝트 사이트에 프로젝트 등록 및 오픈 리포지토리에 데이터 저장</li> <li>• 검색, 분석, 용도 변경, 배포, 표현 방법 지원</li> </ul>
		출판	<ul style="list-style-type: none"> <li>• 카탈로그 작성 및 설명</li> <li>• 배포된 데이터에 대한 일관성 있는 식별 보장</li> <li>• 데이터에 대한 정보 배포</li> <li>• 데이터 공개 프로세스</li> <li>• DOI(Digital Object Identifier)</li> <li>• 데이터 인용</li> <li>• 데이터 배포</li> <li>• 데이터 공유</li> <li>• 저작권 설정</li> <li>• 데이터 홍보</li> </ul>
모든 라이프 사이클 단계	기술 (Describe)	메타데이터 (Metadata)	<ul style="list-style-type: none"> <li>• 관리, 설명, 기술, 구조 및 보존 메타데이터를 할당</li> <li>• 메타데이터에 대한 표준 정의</li> <li>• 메타데이터 생성 및 유지</li> <li>• 메타데이터 작성 도구</li> </ul>
		도큐멘테이션 (Documentation)	<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>
	보장 (Assure)		<ul style="list-style-type: none"> <li>• 데이터의 무결성을 유지하면서 진본, 신뢰성 및 가용성을 보장</li> <li>• 유효성 검사, 허용 가능한 데이터 구조 또는 파일 형식 보장</li> <li>• 품질 보증</li> <li>• 품질 관리</li> <li>• 데이터 품질 문서화</li> </ul>
	백업 및 보안 (Backup & Secure)	백업	<ul style="list-style-type: none"> <li>• 데이터 백업</li> <li>• 관련 표준을 준수하는 안전한 방식으로 데이터 저장</li> <li>• 잠재적인 손실 최소화를 위한 단기적 보존 계획 수립</li> </ul>
보호		<ul style="list-style-type: none"> <li>• 공개 위험 제한 원칙</li> <li>• 이용이 제한된 데이터 수집 방법</li> <li>• 데이터 Enclave</li> <li>• 접근 제한을 통한 개인 정보 보호 및 기밀성 보호</li> <li>• 데이터 익명화</li> <li>• 우발적인 데이터 손실, 손상 및 무단 접근으로부터 데이터를 보호</li> </ul>	

이터를 관리하고 서비스하려는 움직임이 오래 전부터 시도되어 왔다. 또한 해외에서는 이러한 연구 데이터 관리를 위하여 데이터 라이프 사이클을 개념부터 실제 적용까지 모델을 개발하였지만, 국내에서는 아직 많은 연구가 진행

되지 못했다. 이에 따라 본 연구에서는 해외 기관의 데이터 라이프 사이클의 내용을 비교 분석하여 연구 데이터를 관리하기 위한 데이터 라이프 사이클을 제안하였다. 아직 국내에서 연구되지 않은 데이터 라이프 사이클을 제안하

기 위하여 분석된 해외 기관은 DCC, ICPSR, IWGDD, DataONE, USGS 그리고 UK Data Archive 등 6개 기관이다. 6개 기관의 데이터 라이프 사이클을 분석한 결과 공통적으로 도출된 구성 요소는 '계획', '생성 및 수집', '프로세스', '보존', '이관 및 폐기', '접근 및 이용', '기술', '보장' 그리고 '백업 및 보안' 등 9개이다. 도출된 9개의 구성 요소는 향후 국내에서 연구 데이터 관리를 위한 데이터 라이프 사이클을 만들 때 참고할 수 있는 내용으로 해당 요소의 상세 내용은 5장에서 제시하였다.

해외의 경우, 1990년대부터 데이터 보존과 큐레이션을 위하여 데이터 라이프 사이클을 개발하기 시작하였으며 현재는 특정 분야나 커뮤니티의 요구사항을 반영한 연구 데이터 라이프 사이클을 개발하고 있는 추세이지만 국내에서

는 연구가 상당히 미흡한 실정이다. 본 연구에서 개발한 데이터 라이프 사이클 모델은 일반적인 연구 데이터 관리를 위하여 적용될 수 있지만, 특정 분야나 커뮤니티를 지원하기에는 한계가 있다. 이에 따라 추후 학문 분야별 또는 특정 커뮤니티를 지원하기 위한 데이터 라이프 사이클에 대한 연구가 진행될 필요가 있다. 연구 데이터는 학문 분야별 또는 특정 커뮤니티 별로 유형과 성격이 다를 수밖에 없고 해당하는 연구 데이터 관리를 지원하기 위한 특정 요구 조건이 필요할 것이다. 결국 연구 데이터 관리를 위한 데이터 라이프 사이클을 제안하기 위해서는 해당 도메인과 관련된 커뮤니티의 이해 당사자들의 의견을 충분히 수렴한 뒤 해당 연구를 진행해야 할 것이다.

## 참 고 문 헌

- [1] 김주섭, 김선태, 최상기. 2019. 연구 데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건. 『한국문헌정보학회지』, 53(3): 317-344.
- [2] C Jung et al. 2014. "Optimization of data life cycles." *Journal of Physics: Conference Series*, 513: 1-8. [online] [cited 2019. 10. 9.] <doi:10.1088/1742-6596/513/3/032047>
- [3] Data Lifecycle. [online] [cited 2019. 10. 13.] <<https://nmlm.gov/data/thesaurus/data-lifecycle>>
- [4] DataONE Public Participation in Scientific Research Working Group. 2013. *Data Management Guide for Public Participation in Scientific Research*. [online] [cited 2019. 10. 15.] <<https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>>
- [5] Faundeen, J. L. et al. 2013. *The United States Geological Survey Science Data Lifecycle Model*: U.S. Geological Survey Open-File Report 2013-1265, 4 p. <<http://dx.doi.org/10.3133/ofr20131265>>

- [6] ICPSR. 2012. *Guide to Social Science Data Preparation and Archiving*. 5<sup>th</sup> edition. [online] [cited 2019. 10. 14.] <<https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>>
- [7] Interagency Working Group on Digital Data. [online] [cited 2019. 10. 14.] <[https://itlaw.wikia.org/wiki/Interagency\\_Working\\_Group\\_on\\_Digital\\_Data](https://itlaw.wikia.org/wiki/Interagency_Working_Group_on_Digital_Data)>
- [8] Intro to data management. [online] [cited 2019. 10. 17.] <<https://library.cmu.edu/datapub/dms/data/101>>
- [9] IWGDD. 2012. *HARNESSING THE POWER OF DIGITAL DATA FOR SCIENCE AND SOCIETY*. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council January 2009. [online] [cited 2019. 10. 15.] <[https://www.nitrd.gov/About/Harnessing\\_Power\\_Web.pdf](https://www.nitrd.gov/About/Harnessing_Power_Web.pdf)>
- [10] Kristi Miller et al. 2018. Data Management Life Cycle. *Texas A&M Transportation Institute, PRC 17-84F, March 2018* [online] [cited 2019. 10. 12.] <<https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-84-F.pdf>>
- [11] Louise Corti et al. 2014. *Managing and Sharing Research Data: A Guide to Good Practice*. [online] [cited 2019. 10. 13.] <[http://www.sagepub.com/sites/default/files/upm-binaries/61019\\_Corti\\_Managing\\_and\\_sharing\\_research\\_data.pdf](http://www.sagepub.com/sites/default/files/upm-binaries/61019_Corti_Managing_and_sharing_research_data.pdf)>
- [12] Maureen Pennock. 2007. Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information. *Library & Archives*, January 2007. [online] [cited 2019. 10. 8.] <[http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch\\_curation.pdf](http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf)>
- [13] Mohammed El Arass, Iman Tikito, Nissrine Souissi. 2017. Data lifecycles analysis: towards intelligent cycle. *IEEE ISCV, Apr 2017, Fez, Morocco. fhal-01593851f* [online] [cited 2019. 10. 11.] <<https://hal.archives-ouvertes.fr/hal-01593851/document>>
- [14] Open Science. [online] [cited 2019. 10. 12.] <[https://en.wikipedia.org/wiki/Open\\_science](https://en.wikipedia.org/wiki/Open_science)>, <<https://eua.eu/issues/21:open-science.html>>
- [15] Peter Kraker et al. 2011. The case for an open science in technology enhanced learning. *Int. J. Technology Enhanced Learning*, 3(6): 643-654.. [online] [cited 2019. 10. 8.] <[http://www.know-center.tugraz.at/download\\_extern/papers/open\\_science.pdf](http://www.know-center.tugraz.at/download_extern/papers/open_science.pdf)>
- [16] Research data lifecycle. [online] [cited 2019. 10. 14.] <<https://blogs.ntu.edu.sg/lib-datamanagement/data-lifecycle/>>
- [17] Research data management explained. [online] [cited 2019. 10. 12.] <[https://library.leeds.ac.uk/info/14062/research\\_data\\_management/61/research\\_data\\_management\\_explained](https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained)>

- [18] Sarah Higgins, 2008. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, 1(3): 134-140.
- [19] Tanja Wissik and Matej Ďurčo. 2015. Research Data Workflows: From Research Data Lifecycle Models to Institutional Solutions. *CLARIN 2015 Selected Papers · Linköping Electronic Conference Proceedings*, No. 123: 94-107. [online] [cited 2019. 10. 10.]  
<<http://www.ep.liu.se/ecp/123/008/ecp15123008.pdf>>
- [20] VEERLE VAN DEN EYNDEN. 2013. DATA LIFE CYCLE & DATA MANAGEMENT PLANNING. *LOOKING AFTER AND MANAGING YOUR RESEARCH DATA (GOING DIGITAL AND ESRC ATN EVENTS) UK DATA ARCHIVE*, COLCHESTER, 24-25 APRIL 2013. [online] [Cited 2019. 10. 15.]  
<<https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf>>

• 국문 참고자료의 영어 표기

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

- [1] 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.