

연구 데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건*

The Functional Requirements of Core Elements for Research Data Management and Service

김 주 섭 (Juseop Kim)**

김 선 태 (Suntae Kim)***

최 상 기 (Sangki Choi)****

목 차

- | | |
|------------------|--------------------|
| 1. 서 론 | 4. 요소별 기능적 요구사항 도출 |
| 2. 이론적 배경 | 5. RDMS 모델 제안 |
| 3. RDMS 핵심 요소 도출 | 6. 결 론 |

초 록

데이터의 가치 증대, 연구 방법의 패러다임 변화 그리고 오픈 사이언스 등의 구체적 발현은 연구가 더 이상 기존의 학술지와 같은 텍스트 중심이 아닌 데이터 기반으로 전환되고 있음을 나타내고 있다. 본 연구에서는 아직 국내 연구가 미흡한 연구 데이터 관리 및 서비스를 위한 핵심요소와 해당 기능적 요건을 도출하기 위하여 DCC, ICPSR, ANDS 그리고 DataONE에 대한 서비스를 분석하였다. 분석 결과 도출된 핵심 요소는 'DMP 작성지원', '데이터 기술', '데이터 저장', '데이터 공유 및 접근', '데이터 인용' 그리고 '데이터 관리 교육' 등이다. 또한 도출된 핵심 요소에 기능적 요건을 제시함으로써 향후 실질적으로 RDM 서비스를 구축 및 운영하고자 할 때 본 연구의 내용을 적용할 수 있을 것이다.

ABSTRACT

Increasing the value of data, paradigm shifts in research methods, and specific manifestations of open science indicate that research is no longer text-centric, but data-driven. In this study, we analyzed the services for DCC, ICPSR, ANDS and DataONE to derive key elements and functional requirements for research data management and services that are still insufficient in domestic research. Key factors derived include DMP writing support, data description, data storage, data sharing and access, data citations, and data management training. In addition, by presenting functional requirements to the derived key elements, this study can be applied to construct and operate RDM service in the future.

키워드: 연구 데이터, 연구데이터 관리 서비스, 오픈 사이언스, 리퍼지토리, 데이터 관리 계획

Research Data, RDMS (Research Data Management Service), Open Science, Repository,
DMP (Data Management Plan)

* 본 연구는 2019년도 전북대학교 연구기반 조성비 지원을 받아 수행되었음.

** 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(choisk@jbnu.ac.kr / ISNI 0000 0004 6468 1351) (공동저자)

논문접수일자: 2019년 7월 21일 최초심사일자: 2019년 8월 6일 게재확정일자: 2019년 8월 16일
한국문헌정보학회지, 53(3): 317-344, 2019. [http://dx.doi.org/10.4275/KSLJIS.2019.53.3.317]

1. 서론

1.1 연구의 배경 및 목적

하드웨어 발전과 고성능 네트워크 그리고 센서 장비의 확산은 많은 연구 분야에서 다양한 데이터를 만들어내고 있다(김선태, 이태영, 김용 2012). 작은 센서로부터 거대 실험 장치들이 원유라 불리는 데이터를 실시간으로 생산함에 따라 데이터양은 기하급수적으로 증가하게 되었으며 이러한 데이터를 체계적으로 관리하기 위한 서비스가 각 연구 분야별로 요구되고 있는 실정이다. 최근, 연구 패러다임과 데이터 인식의 변화에 따라 법적, 제도적, 기술적 장벽 없이, 모든 이용자에게 과학적 연구 결과물에 대한 접근이 가능하도록 하는 오픈 사이언스(Open Science) 운동이 전 세계적으로 활발하게 전개되고 있다. 이러한 오픈 사이언스 운동의 일환으로 국내에서도 국민세금으로 생산된 연구결과물이 국가 자산, 기관 자산으로서 재사용될 수 있는 환경에서 체계적으로 관리 및 보존되어야 한다는 인식이 확산되고 있다. 오픈 사이언스는 일반적으로 공공 기금 연구의 산출물을 과학 커뮤니티, 비즈니스 부문 또는 사회에 디지털 형식으로 보다 널리 접근할 수 있도록 만드는 노력을 말하며, 2015년에 OECD의 “Making Open Science a Reality” 보고서를 통해 오픈 사이언스에 관한 논의의 장이 확장되었다(OECD 2015). 이 오픈 사이언스는 오픈 액세스(Open Access), 오픈 데이터(Open Data), 오픈 리포지토리(Open Repository) 등으로 구체화되는데, 여기에서 데이터가 핵심의 위치에 있다. 특히, 여기에서 언급한 오픈 데이

터(Open Data)는 “모든 사람이 자유롭게 사용, 재사용 및 재배포 할 수 있는 데이터(OPEN DATA HANDBOOK 2012)”로 학술 연구 영역에서는 이러한 데이터를 ‘연구 데이터’라 부르고 있다. 연구 데이터는 많은 시간과 돈을 들여 만든 중요한 정보자원으로서 결과에 대한 검증, 연구의 효율성, 데이터 보안 및 데이터 손실 최소화, 데이터의 폭넓은 보급과 영향력 증대, 데이터의 재사용을 통한 연구의 연속성 가능 그리고 편당 기관의 요구 사항을 준수하기 위해 적절하게 관리가 되어야 한다(UCD Library 2019).

데이터에 대한 가치 증대와 연구 패러다임의 변화 그리고 오픈 사이언스 운동 확산은 연구의 방식이 기존 문헌에서 데이터로 이동하고 있다는 것을 보여주고 있다. Jim Grey는 대규모 과학 데이터셋과 관련된 데이터 엔지니어링 문제에 접근하는 방법을 체계화하는 법칙을 공식화하였는데, 해당 법칙의 첫 번째는 ‘과학 컴퓨팅은 점점 더 데이터 집약적이 되어 가고 있다’라는 것이다(Tony Hey, Stewart Tansley and Kristin Tolle 2009). 2016년 OECD는 ‘OECD Science, Technology and Innovation Outlook 2016’라는 보고서를 통해 가설과 ‘대이론(grand theories)’ 개발을 통한 전통적인 접근 방식은 방대한 양의 데이터로 시작하는 데이터 중심의 연구로 보완될 것이며 이 방향으로의 변화는 이미 명백하다고 기술하였다(OECD 2016). 이것은 연구의 행태가 기존의 실험이나 이론 중심에서 데이터 중심으로 옮겨가고 있다는 것을 보여주고 있다. 이러한 연구 패러다임의 변화는 전 세계적으로 데이터에 대한 인식의 변화를 가져오게 되었다.

연구 공공 자금으로 연구된 산출물에 대한 공유와 재사용이 확산되고 있는 가운데 미국, 영국, 호주 등 각 국가별로 연방기관이나 연구회를 중심으로 연구 데이터의 공유 및 활용 정책을 시행하고 인프라를 구축 및 운영하고 있다(이상환 2019). 또한, 2013년을 시작으로 미국, 영국, EU 그리고 일본 등은 공공 연구를 통해 산출된 연구 데이터의 선제 관리 및 접근 확대 지침을 제정하였다. 하지만, 한국은 국가 R&D 사업을 통해 산출되는 연구데이터는 증가하고 있지만 이러한 연구 데이터 관리를 위한 규정이 미흡한 실정이다.¹⁾ 실제 안형준 외(2017)는 연구 현장에서 데이터 관리 현황을 조사한 결과 개인이 관리하는 경우가 43.8%로 다수를 차지하고 있으며 관리 실태도 표준 절차에 따라 관리되기 보다는 개인이 데이터를 생산한 후 저장이나 보관 등 단순관리에 그치고 있다고 분석하였다. 이것은 연구 데이터가 기관 중심이 아닌 개인적으로 관리가 되고 있고 기관의 연구 데이터 제도 등 관리 기반이 부족하다는 것을 나타낸다. 본 연구에서는 국내에서 수집 및 생산되는 연구 데이터에 대한 체계적인 관리를 위하여 해외 선진기관에서 운영 중인 연구 데이터 관리 및 서비스를 분석하여 해당 서비스를 위한 핵심 요소의 기능적 요건을 도출하는 것을 목적으로 하였다. 이를 위해 데이터 큐레이션 센터인 DCC, 사회과학 분야 데이터 아카이브인 ICPSR, 호주 국가 데이터 서비스인 ANDS, 마지막으로 지구를 위한 데이터 관측 네트워크인 DataONE 등 4개의 선진 기관을 분석 대상으로 하였다.

1.2 연구 방법과 범위

본 연구에서는 연구 데이터 서비스 및 관리를 위한 모델을 제안하기 위하여 해외 선진 기관의 서비스를 분석하였다. 본 연구에서 분석한 해외 선진 기관은 다음과 같다.

- DCC(Digital Curation Centre)
- ICPSR(Inter-university Consortium for Political and Social Research)
- ANDS(Australian National Data Service)
- DataONE(Data Observation Network for Earth)

해외의 모든 선진기관에서 운영 중인 연구 데이터 관리 및 서비스를 조사하여 구성 요소를 도출하는데 한계가 있기 때문에 위에서 제시하는 4개의 기관을 선택하여 연구의 범위를 제한하고자 하였다. 먼저, DCC를 선택한 이유는 DCC가 연구 데이터 관리에 관한 정책을 제공하는 기관으로 미국 등 많은 선진국에서 연구 데이터 관리(Research Data Management, 이하 RDM) 서비스를 개발하기 위하여 참조하는 기관이기 때문이다. ICPSR은 실제로 사회과학 분야의 데이터 서비스를 하고 있는 곳으로 데이터 관리 및 큐레이션 서비스를 제공하고 있으며 해당 분야에 특화된 데이터 리포지토리로서 역사적으로 오래된 기관이다. 다음으로 ANDS는 다분야 학문을 위한 연구 데이터 관리 서비스로 국가가 주도적으로 운영하고 있어 이번 분석 기관에 포함하였다. 마지막으

1) 현재, 연구 데이터 정의와 데이터 관리 계획 도입 및 DMP 정보 등록 근거가 신설되었다(2019.03.19. 개정).

로 DataONE은 다 분야 과학 데이터의 보존, 액세스, 이용 및 재사용 그리고 데이터 아카이빙을 지원하는 곳으로 전 세계적으로 성공적인 데이터 인프라로 인정받은 곳으로 선정하게 되었다.

제시된 연구 목적을 달성하기 위하여 위의 4가지 선진기관에서 제공하는 연구 데이터 관리 및 서비스를 분석하여 공통된 구성 요소를 도출하였다. 또한 도출된 공통 요소의 내용을 4개의 기관을 중심으로 비교하여 최종적으로 제시한 모델의 내용에 제안하였다. 공통된 요소를 도출하기 위하여 본 논문에서는 김지현(2014)의 선행 연구에서 제시한 RDM 구성 요소를 분석 기준으로 활용하였다. <그림 1>은 본 연구에서 수행한 연구 데이터 관리 및 서비스 모델을 제안하기 위한 연구 절차를 나타낸 것이다.

2. 이론적 배경

전 세계적으로 국가의 공적 자금을 통해 생산된 연구 성과물에 대하여 일반 대중에게 공개하려는 움직임이 활발하게 전개되고 있다. UN, OECD 등 국제기구뿐만 아니라 미국의 NSF, 영국의 UKRI 등 국가의 연구기관들도 자국의 공적 자금으로 생산된 연구 성과물을 온라인에 공개하도록 정책을 마련하였다(Crouzier 2015). 주요 선진국의 연구 데이터 관리 동향을 살펴 보면, <표 1>과 같다.

최근에는 연구자들에 의해 생산된 연구 성과물 즉, 논문이나 연구보고서뿐만 아니라 연구 시작부터 연구 종료될 때까지 연구자에 의해 생산된 연구 데이터의 관리 및 공유에 대한 논의가 활발하게 이루어지고 있다. 연구 성과물에서부터 연구 데이터에 대한 일반 대중으로

5단계	연구 데이터 관리 및 서비스를 위한 모델 제안
4단계	도출된 공통 구성 요소의 내용 비교 분석 - DCC, ICPSR, ANDS, DataONE
3단계	연구 데이터 관리 및 서비스를 위한 공통 구성 요소 도출
2단계	해외 선진 기관의 연구 데이터 서비스 사례 분석 - DCC, ICPSR, ANDS, DataONE
1단계	연구 방향 및 범위 설정을 위한 브레인 스토밍

<그림 1> 연구 데이터 관리 및 서비스 모델 개발 절차

<표 1> 주요 선진국의 연구 데이터 관리 동향(김선태 2015)

주요 선진국	동향
미국, 독일	연구데이터의 글로벌 유통체계 선도(NASA GCMD, TIB DataCite 등)
영국, 유로	연구데이터 보존 및 공유, 연구자 분석 및 교육(JISC, DCC, PARSE Insight)
호주, 중국	연구데이터를 위한 체계적인 국가적 인프라구축(ANDS, ARCS, 중국 NSTI)
캐나다, 호주	자국의 연구데이터 수집 및 게이트웨이 서비스(CISTI, ANDS)

공개는 오픈 사이언스의 실천 운동으로 전 세계로 확산되고 있는 추세다(예상준, 장호, 김선태 2019). 이번 장에서는 오픈 사이언스에 대한 일반적인 함의와 각 국가별 정책에 대하여 정리하고 연구 데이터 관리 및 서비스와 관련된 선행 연구를 제시하였다.

2.1 오픈 사이언스

전 세계적으로 법적, 제도적, 기술적 장벽 없이, 모든 이용자에게 과학적 연구 결과물에 대한 접근이 가능하도록 하는 운동으로 오픈 사이언스(Open Science) 운동이 활발하게 전개되고 있다. 오픈 사이언스에 대한 최초의 용어는 1998년 Steve Mann에 의해 만들어졌으며(Wikipedia 2019) 2004년에는 OECD 과학기술장관 회의에서 연구데이터에 대한 접근성 제고를 위한 노력이 시작되었다(OECD 2007). 또한, OECD는 2015년에 “MAKING OPEN SCIENCE A REALITY”라는 보고서를 통해 오픈 사이언스에 대한 용어정의를 다음과 같이 제시하였다.

오픈 사이언스란 연구를 가속화하기 위한 수단으로서 공개적으로 자금을 지원하는 연구 결과(출판물 및 연구 데이터)의 주요 산출물을 디지털 형식으로 접근 가능하도록 만드는 연구자, 정부, 연구비 지원 기관 또는 과학 커뮤니티 자체의 노력을 의미한다(OECD 2015, 7).

Fecher와 Friesike(2013)은 문헌 검토를 토대로, 다섯 가지 오픈 사이언스 스쿨(School)을 제안함으로써 다음과 같은 전반적인 담론을 구성하였다.

- 인프라 스쿨(infrastructure school): 기술 아키텍처(technological architecture)
- 공공 스쿨(public school): 지식 생성의 접근성
- 측정 스쿨(measurement school): 대체 가능한 영향 측정(alternative impact measurement)
- 민주적 스쿨(democratic school): 지식 접근
- 실용적 스쿨(pragmatic school): 공동 연구

Fecher와 Friesike(2013)는 과학 연구의 모든 과정을 일반 대중에게 접근 가능하도록 개방되어야 함을 담론으로 제시하면서 5가지의 핵심 가치에 대한 해결 방법으로 오픈 액세스, 오픈 데이터, 협력 작업을 위한 플랫폼 등을 제안하고 있다.

오픈 사이언스의 개념이 전 세계적으로 공유되면서 각 국가별로 오픈 사이언스 정책의 범위도 넓어지고 있다. 즉, 2000년 전후로 오픈 사이언스 정책은 오픈 액세스를 가리키고 그 이후로는 공공 영역을 중심으로 공적 자금으로 수행된 연구 논문에 대한 공개토록 하는 정책으로 확대되었다(신은정 외 2017). <표 2>는 EU를 비롯한 주요국의 오픈 사이언스 정책을 정리한 것이다.

2.2 선행 연구

김지현(2012)은 대학 내 연구자들의 연구 데이터 관리 현황에 관하여 연구한 결과, 생산된 데이터는 개인 PC나 이동식 매체에 저장하고 있고 데이터의 수집은 다수가 기존 논문에서 추출하거나 개인적인 연락을 통해 이루어지

〈표 2〉 주요국의 오픈 사이언스 정책 동향(신은정 외 2017, 32-52)

주요 국가		오픈 사이언스 정책 동향
유럽 연합(EU)		<ul style="list-style-type: none"> • 2016년, 오픈 사이언스 정책 플랫폼(Policy Platform)을 구성하고 관련 전문가들의 정책제언을 수렴 • 2016년, 유럽차원의 오픈 사이언스 추진동향을 모니터링하기 위한 모니터링 프레임워크(Monitoring Framework)를 만드는 작업에 착수 • 2016년부터 유럽 지역 내 데이터 공유 및 연구협력을 지원하는 유럽 오픈 사이언스 클라우드(European Open Science Cloud)를 구축, 운영하기 위한 작업에 착수
핀란드		<ul style="list-style-type: none"> • 핀란드 학술원(Academy of Finland), TESKE와 같은 공공 연구지원 기관들이 점진적으로 오픈 액세스 정책을 도입 · 이행해 나가고 있음 • 2017년, 오픈 액세스(open access)를 넘어 오픈사이언스 전반에 관한 국가계획인 『National Plan Open Science』를 발표
네덜란드		<ul style="list-style-type: none"> • 2017년, 오픈 액세스를 넘어 오픈 사이언스 전반에 관한 국가계획인 『National Plan Open Science』를 발표 • 2017년, 오픈 사이언스 정책의 원활한 이행을 위해 오픈 사이언스 국가플랫폼(OCW)을 구축 · 운영
미국	연방정부	<ul style="list-style-type: none"> • 연방정부 지원으로 생산된 연구 성과에 대한 접근성 제고를 위한 OSTP 메모 <ul style="list-style-type: none"> - 출판물에 대한 접근성을 제고하려는 오픈 액세스 정책 - 연구 데이터에 관한 접근성을 제고하려는 오픈 데이터 정책
	NIH	<ul style="list-style-type: none"> • 2005년 『공공 액세스 정책(Public Access Policy)』을 발표
일본		<ul style="list-style-type: none"> • 2015년 3월 『일본 오픈 사이언스 활성화 방안(Promoting Open Science In Japan: Opening Up a New Era for the Advancement of Science)』이 발표 • 2017년 4월, 공공연구에서 생산된 연구 데이터에 대해서 데이터 관리 계획(DMP)수립을 권고하는 정책 발표

고 있다고 밝혔다. 연구자들의 데이터 공유는 조직 내 연구자들과 이루어지거나 외부에서 요청시 제공해준다고 하였다. 이러한 연구데이터의 관리체계에 대한 부실에 따라 저자는 연구데이터 아카이빙 서비스 개발의 필요성을 주장하였다.

김지현(2014)은 미국의 연구중심대학도서관에서 운영하고 있는 RDM 서비스 구성요소와 웹사이트의 내용을 9가지의 기준으로 분류하였다. 조사 대상의 대학에서 RDM 서비스를 구축하여 운영하고 있으며, 메타데이터 제시, 리포지터리 안내 그리고 교육 및 컨설팅 서비스를 제공하고 있다고 제시하였다. 특히, 국내 서비스 도입을 위해서는 기존 도서관 조직과 다양한 기관과의 협력, 서비스 운영을 위한 사서의 지식과 역량 및 교육 프로그램이 필요하다는

시사점을 도출하였다.

심원식(2016)은 미국 대학 중 활발하게 연구데이터 서비스를 수행중인 4개 대학(Virginia, Johns Hopkins, Purdue, Illinois)의 사례를 비교 분석하여 전담조직의 구성, 교육 서비스, 컨설팅 서비스 그리고 시스템 인프라의 관점에서 시사점을 도출하였다. 저자는 국내 대학에 연구데이터 서비스를 적용할 수 있는 영역으로 이용자 교육을 제시하였다. 대학원생들의 연구데이터 관리와 공유를 위한 이용자 교육을 실시함으로써 체계적인 관리가 가능하다고 보았으며 또한, 대학도서관 사서들이 연구데이터의 관리 및 공유에 관한 원칙과 기법에 대해 학습하는 것이 필요하다고 연구에서 제시하고 있다.

최명석 등(2017)은 국가차원의 데이터 공

유 활용 정책 수립을 위해 국내 과학기술 분야 연구기관의 과학데이터(연구 데이터) 관리·활용 현황을 살펴보고 데이터 관리 계획(Data Management Plan, 이하 DMP) 수립 의무화 정책추진, 과학데이터 성과 인정 및 인센티브 방안 마련, 개인정보보호 등 데이터 활용 규제 완화 방안 마련, 과학데이터 품질 관리 체계 수립, 국가차원의 과학데이터 관리·공유 인프라 구축, 과학데이터 공유·활용 인식 제고, 역할 정의 및 전문 지원 기관 지정 등 7가지의 구체적인 개선방향을 도출하였다.

박미영 등(2017)은 국내외 RDMS(Research Data Management Service, 이하 RDMS)의 하위 요소를 비교 분석하여 DMP, 연구데이터 문서화, 데이터 기술(포맷), 데이터 저장, 데이터 공유 및 접근(사용) 그리고 데이터 보존 등 5가지 단계로 구분하여 제시하였다. 특히, 생명공학분야 연구데이터의 공유사례를 리파지토리 유형별로 분석하였는데, 기관 리파지토리 기탁형, 데이터뱅크 기탁형, 전문데이터센터 기탁형 그리고 데이터 아카이브 기탁형 등 4가지로 구분하였다.

Yu(2017)는 연구데이터 서비스를 제공하기 위한 도서관의 역할에서 문헌과 ARL 그리고 ACRL로부터 수집된 사례를 기반으로 DMP, 메타데이터 컨설팅과 도구 제공, 데이터 공유와 접근, 기관 저장소 제공 그리고 데이터 아카이빙 등 5가지 분야를 도출하였다. 또한, 저자는 연구자들의 데이터 보관 및 공유를 위해 로컬 저장소와 같은 오픈 액세스 저장소를 활용하도록 도서관이 안내하고 있으며, RDS(Research Data Service, 이하 RDS) 정책 및 인프라 개발에 대한 논의는 충분하지 않거나 거의 존재하

지 않는다고 기술하고 있다.

Perrier와 Barnes(2018)는 연구자를 위한 연구 데이터 관리 서비스 개발이라는 논문에서 포커스 그룹으로부터 원활한 인프라, 데이터 보안, 기술 및 지식 개발 그리고 데이터 공개에 대한 불안 등 4가지의 테마를 확인하였다. 특히, 저자는 도서관이 연구자들의 연구데이터 관리를 지원하기 위한 우선 순위를 다음과 같이 데이터 관리와 연구 지원 서비스 2가지 분야로 구분하여 나타냈다(〈표 3〉 참조).

〈표 3〉 Perrier와 Barnes의 연구데이터 관리를 지원하기 위한 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 리파지토리 확인
연구 지원 서비스	(1) 자금 지원 기관 요구 사항
	(2) 저널 요구 사항
	(3) 지적재산권
	(4) 데이터셋 인용
	(5) 리파지토리에서 데이터 찾는 방법
	(6) 동의서
	(7) 기밀 유지

Cox(2017) 등은 「대학도서관 연구데이터 관리 개발」이라는 논문에서 호주, 캐나다, 독일, 아일랜드, 네덜란드, 뉴질랜드 및 영국 등 7개 나라의 고등교육 도서관의 설문조사 결과를 제시하였다. 설문조사 결과, 기술 서비스(데이터 목록 제공 및 데이터 큐레이션)보다는 자문 및 컨설팅 서비스(DMP 작성 지원 및 데이터 관련 교육)에 초점을 두고 있고 정책 및 거버넌스, 서비스 개발 및 직원 배치 그리고 기술 등은

많은 국가에서 RDM 개발의 주요 과제로 확인되었다.

Steeleworthy(2014)는 『연구데이터 관리와 캐나다 대학 도서관』이라는 연구 논문에서 RDM 서비스가 성공하기 위해서는 리에중 사서, 수서 전문가, 정책 전문가, IT 전문가, 아키비스트 및 보존 전문가 등 기존 도서관 조직의 기능과 더불어 캠퍼스 이해관계자들과의 협업, 기술 및 정책 분석 기술 그리고 프로젝트 관리 통찰력이 필요하다고 주장하였다. 또한, 도서관이 연구원, 연구 윤리위원회, 연구 기관 및 컨소시엄의 도서관과 같은 이해관계자와 파트너 관계를 맺음으로써 견고한 서비스를 구축할 수 있고 이러한 도서관과 교내 기관과의 협력이 서비스를 성공할 수 있게 만든다고 기술하였다.

Koltay(2017)는 『대학과 연구 도서관의 Research 2.0과 연구 데이터 서비스』라는 논문에서 연구 데이터 관리, 데이터 큐레이션, 데이터 관리 체계 및 데이터 리터러시 교육으로 구성된 RDS에 대하여 2014년부터 2016년까지의 출판된 문헌을 검토하였다. Research 2.0이 나라와 일부 지역 환경에서 다르게 발전하고 있지만 도서관 및 사서의 도움이 필요하며 데이터 관리 계획 개발 지원 또는 표준화 된 메타데이터 작성 등 전통적인 서비스는 여전히 유효하다고 밝히고 있다. 특히, 다른 학술 기관에서 다루지 않는 서비스를 발견하도록 도서관 및 사서들이 노력해야 한다고 주장하였다.

국내의 선행 연구와 본 연구의 차이점은 선행 연구의 대부분이 RDM 서비스의 대표적인 구성 요소를 도출하거나 또는 우선 수행해야 할 서비스를 제시하는 것으로 그쳤지만, 본 연

구에서는 도출된 구성 요소에 한하여 상세 항목을 제시함으로써 실질적으로 RDM 서비스를 구축과 운영하고자 할 때 본 연구 내용을 실무에 활용할 수 있다는 점이다.

3. RDMS 핵심 요소 도출

이번 장에서는 DCC, ICPSR, ANDS 그리고 DataONE 등 연구 데이터 관리 및 서비스를 위한 내용을 분석 및 비교하여 공통된 구성 요소를 도출하고자 한다.

3.1 DCC(Data Curation Center) 요소 분석

DCC는 영국 정부기관인 JISC(Joint Information Systems Committee) 산하의 기관으로, 연구 데이터 관리를 위한 기술과 능력을 양성하는데 중점을 둔 국제적으로 인정받은 디지털 큐레이션(digital curation) 전문 센터다. DCC의 주요 역할은 다음과 같다(DCC 2019).

- 디지털 연구데이터를 저장, 관리, 보호 및 공유하려는 연구 조직에 전문적인 조언과 실질적인 도움을 제공
- How-to 가이드, 사례 연구 및 온라인 서비스를 비롯한 다양한 정보자원에 대한 접근 제공
- 연구자 및 데이터 관리자에게 데이터를 효과적으로 관리하고 공유하는데 필요한 기술을 갖추기 위한 교육 제공
- 정책 개발 및 DMP와 같은 이슈에 대한 컨설팅 및 지원 제공

〈표 4〉는 영국 DCC에서 제공하는 RDM 서비스의 구성요소와 내용을 정리한 것이다. DCC의 RDM 구성요소는 ‘RDM 정책과 전략’, ‘업무계획과 지속성’, ‘가이드, 교육 및 지원’, ‘DMP’, ‘현용 데이터 관리’, ‘데이터 선정 및 수집’, ‘데

이터 리파지토리’ 그리고 ‘데이터 기술’ 등 8개로 구성되어 있다.

먼저, ‘RDM 정책과 전략’은 연구 데이터 관리를 위한 전략 및 정책을 개발하고 그리고 데이터 정책 및 전략을 조정하는 역할을 수행한

〈표 4〉 영국 DCC의 RDM 서비스 구성요소(Jones, Pryor and Whyte 2013)

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

다. 두 번째, '업무계획과 지속성'은 중장기 사업 계획의 개발, 사업 계획에 따른 예산 비용 산출 그리고 지속 가능성 이슈 및 비용 문제 해결을 위한 서비스를 제공한다. 세 번째, '가이드, 교육 및 지원'은 연구자를 지원하고 지원 인력을 재교육하기 위한 RDM 교육과 심층적인 컨설팅 서비스를 제시한다. 네 번째, 'DMP'는 DMP 서비스에 대한 필요성, DMP 템플릿 및 가이드, DMP 도구, 교육과 컨설팅 그리고 DMP 작성을 누가 담당할 것인지에 대한 것을 제안한다. 다섯 번째, '현용 데이터 관리'는 데이터 저장을 위해 연구 데이터 스토리지 할당 및 관리 절차 개발, 다양한 범위의 연구 상황과 요구 사항을 충족하는 데이터의 생성, 관리 및 공유를 지원하는 유연한 RDM 시스템 제공을 제시한다. 여섯 번째, '데이터 선택 및 수집'은 연구자가 데이터를 선택하고 기탁하도록 권장하며 전공별 가이드, 기탁 도구 안내 및 동의서 작성 그리고 선택 결정을 위한 기준을 제공한다. 일곱 번째, '데이터 리포지토리(Data Repository)'는 필요한 경우 기관 리포지토리 개발(Institutional Data Repository) 및 유지, RDM 전략에서 외부 데이터 리포지토리가 수행할 역할 결정 그리고 관련 서비스 안내 및 지원을 제시한다. 마지막으로 '데이터 기술'은 연구 데이터셋을 기록하는데 필요한 메타데이터를 선정하고 적용하도록 제안하고 있다.

3.2 ICPSR(Inter-university Consortium for Political and Social Research) 요소 분석

정치 및 사회 연구를 위한 대학 간 컨소시엄인 ICPSR은 1962년에 설립되었으며 특히, 사

회과학 연구 인프라의 필수 부분으로서 연구 및 교육을 위한 방대한 사회과학 데이터 아카이브를 유지 관리하고 있다(8,000개 이상의 개별 연구/65,000개 이상의 데이터 세트를 가진 설문 조사, Wikipedia 2019). 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(Data Management & Curation) 서비스를 6가지 카테고리로 범주화하여 제시하고 있다. 6가지는 품질, 보존, 접근, 기밀유지, 인용, 도구 및 서비스로 구성되어 있으며 각각의 내용은 <표 5>와 같다.

첫 번째, '품질(Quality)'은 조직적, 절차적 및 기술적으로 신뢰할 수 있는 데이터를 위해 데이터 공유 및 아카이빙의 중요성, 데이터 아카이빙 및 보존을 위한 사전 계획 그리고 데이터 라이프 사이클(Data Life Cycle)을 제공하고 있다. 두 번째, '보존'은 모든 조직이 자체 정책 프레임 워크를 개발하는 데 사용할 수 있는 디지털 보존 정책 프레임 워크에 대한 모델을 제안한다. 세 번째, '기밀 유지'는 데이터의 접근과 배포에 대한 것으로 책임 있는 데이터 사용, 데이터 재배포 검토 기준 그리고 자연 배포와 관련된 내용으로 구성되어 있다. 네 번째, '기밀유지'는 각 연구의 공개 위험을 평가하는 것으로 공개 위험, 식별자, 기밀 유지와 기관 검토 위원회, 제한된 데이터 액세스 수준, 컨설팅을 제공한다. 다섯 번째, '데이터 인용 방법'에 관한 것으로 저자, 타이틀, 배포자, 날짜, 버전, 영구 식별자에 대한 내용을 제안한다. 마지막으로 '도구 및 서비스'는 기밀유지(공개 검토,

〈표 5〉 ICPSR Data Management & Curation의 하위요소와 내용(ICPSR Data Management & Curation 2019)

구성요소	내용
품질(Quality)	<ul style="list-style-type: none"> • 아카이빙을 위한 데이터 준비 • 리파지토리 • ICPSR에서의 데이터 큐레이팅(data curating)
보존(Preservation)	<ul style="list-style-type: none"> • 신뢰할 수 있는 디지털 리파지토리 • ICPSR의 디지털 보존 정책 및 계획
접근(Access)	<ul style="list-style-type: none"> • 접근과 배포
기밀유지(Confidentiality)	<ul style="list-style-type: none"> • 기밀유지에 대한 ICPSR의 접근법
인용(Citation)	<ul style="list-style-type: none"> • 데이터 인용 방법(How to Cite Data)
도구 및 서비스(Tools & Services)	<ul style="list-style-type: none"> • 기밀유지(Confidentiality) • 이용-제한 데이터(Restricted-use data) • 데이터 프로세싱(Data Processing) • DMP

텍스트 익명 도우미 도구), 이용-제한 데이터 (물리적 데이터 고립, 가상 데이터 고립), 데이터 프로세싱(Nesstar Publisher, Open Science Framework, OpenRefine) 그리고 데이터 관리 정책(DMP Tool)에 관한 도구와 서비스로 구성되어 있다.

3.3 ANDS(Australia National Data Service) 요소 분석

호주 국립 데이터 서비스(ANDS)는 호주 국립 대학교(ANU) 및 연방 과학 산업 연구 기구(CSIRO)와 협력하여 Monash 대학이 이끄는 파트너십으로 NCRIS (National Collaborative Research Infrastructure Strategy)를 통해 자금을 지원받고 있다. ANDS는 2008년에 공식적으로 설립된 이래로 국제 연구 데이터 커뮤니티에서 중요한 역할을 담당해 왔으며 호주의 연구 기관, 정부 기관 및 문화 기관의 연구 데이터를 찾고 접근하고 재사용 할 수 있는 Research Data Australia 포털이 핵심 서비스이다(ANDS

About us 2019). ANDS는 2013년 1월에 「연구 데이터 관리 실재」라는 가이드에서 연구 데이터 관리 단계별로 연구 데이터 큐레이션을 제시하였다. 〈표 6〉은 ANDS의 연구 데이터 관리 단계를 나타낸 것이다.

ANDS의 연구 데이터 관리 단계는 ‘연구 전(Pre Research)’, ‘연구 중(During Research)’ 그리고 ‘연구 후(During Research)’ 세 가지 단계로 구성된다.

첫 번째, ‘연구 전’ 단계는 프로젝트 시작 전을 의미하며 해당 세부 항목은 ‘연구 데이터 관리 프레임워크’, ‘위험 관리 계획(Risk Management Plan)’, ‘연구 데이터 관리 계획(Research Data Management Plan)’, ‘윤리적 허용(Ethical Clearance)’ 그리고 ‘교육 및 인도(Training and Induction)’ 등을 포함한다. ‘연구 데이터 관리 프레임워크’는 정책과 절차, IT 인프라, 지원 서비스 그리고 메타데이터 관리에 대한 항목이고, ‘위험 관리 계획’은 식별, 접근, 완화 그리고 모니터링 등 연구 도중 발생할 수 있는 데이터 손실 또는 개인 정보 보호, 기밀 유지 또는 저작권

〈표 6〉 ANDS의 연구 데이터 관리 단계별 항목(MERCURY PROJECT SOLUTIONS 2013)

단계	내용
연구 전 (PRE-RESEARCH)	• 연구 데이터 관리 프레임워크
	• 위험 관리 계획(Risk Management Plan)
	• 연구 데이터 관리 계획(Research Data Management Plan)
	• 윤리적 허용(Ethical Clearance)
	• 교육 및 인도(Training and Induction)
연구 중 (During Research)	• 정책 준수 모니터링(Policy Compliance Monitoring)
	• 리스크 모니터링 및 커뮤니케이션(Risk Monitoring and Communication)
	• 연구 데이터 수집 및 분석
	• 메타데이터 생성
	• 저장 및 접근
연구 종료 후 (Post Research)	• 연구 데이터 퍼블리싱
	• 연구 데이터 등록
	• 지속적인 큐레이션(Ongoing Curation)
	• 사용 모니터링(Usage Monitoring)

위반에 관련한 단계별 조치 사항이다. ‘DMP’는 데이터 정리, 메타데이터, 백업, 아카이빙, 버전 관리, 공유, 보안 그리고 거버넌스 등 데이터 관리 계획을 제시한다. 다음으로 ‘윤리적 허용’은 동의, 익명, 접근 통제 그리고 라이선싱에 대한 항목이며 ‘교육 및 인도’는 모든 직원을 대상으로 하는 교육과 가이드 및 보충교육을 말한다.

두 번째, ‘연구 중’ 단계는 프로젝트 동안을 의미하며, 해당 세부 항목은 ‘정책 준수 모니터링’, ‘리스크 모니터링 및 커뮤니케이션’, ‘연구 데이터 수집 및 분석’, ‘메타데이터 생성’ 그리고 ‘저장 및 접근’ 등으로 구성된다. ‘정책 준수 모니터링’은 검토 또는 감사 프레임워크를 통해 준수되지 않은 부분을 조기에 파악하고 시정 조치에 대한 것을 포함한다. ‘리스크 모니터링 및 커뮤니케이션’은 모니터링과 검토를 통한 위험요소 확인 및 우선순위 재평가를 포함하며, ‘연구 데이터 수집 및 분석’은 동료와의 데이터 공유, 데이터 파생 그리고 큐레이션에

대한 항목이다. ‘메타데이터 생성’은 수집과 개체 레벨에서 풍부한 메타데이터 생성을 의미하며, ‘저장 및 접근’은 데이터에 대한 충분하고 적절한 저장 시설 확보와 데이터의 재사용과 공유에 대한 항목이다.

마지막 ‘연구 후’ 단계는 ‘연구 데이터 퍼블리싱’, ‘연구 데이터 등록’, ‘지속적인 큐레이션(Ongoing Curation)’ 그리고 ‘사용 모니터링(Usage Monitoring)’ 등을 세부항목으로 제시했다. ‘연구 데이터 퍼블리싱’은 잘 기술되고 인용가능하며 탐색가능한 그리고 재사용할 수 있는 연구 데이터 출판에 대한 항목이며, ‘연구 데이터 등록’은 RDA(Research Data Australia)를 통한 등록 권고, DOI 등을 통한 연구 데이터 재사용 및 연구 효과 측정에 대한 내용이다. ‘지속적인 큐레이션’은 공유되지 않거나 불필요한 연구 데이터에 대한 보관 및 폐기와 관련된 항목이며, ‘사용 모니터링’은 온라인 도구를 이용한 인용 추적과 관련된 내용이다.

3.4 DataONE(Data Observation Network for Earth) 요소 분석

2009년 최초 DataNet 프로그램 중 하나로써 미국 국립 과학 재단(US National Science Foundation)에 의해 지원된 DataONE은 지구 관측 데이터에 대한 접근하기 위한 환경 및 생태학을 위한 플랫폼이다. DataONE은 주요 사이버인프라의 구축과 교육 및 아웃리치 프로그램을 통한 다 분야 과학 데이터의 보존, 액세스, 이용 및 재사용을 지원하며, 과학자들이 생산한 생태 및 환경데이터에 대한 과학 데이터 아카이빙을 제공한다(DataONE 2019). DCC, ICPSR 그리고 ANDS와 다르게 DataONE은 '연구 데이터 관리' 또는 '데이터 큐레이션'이라는 명칭으로 서비스 항목을 제공하고 있지 않지만, 2013년 2월에 해당 플랫폼 워킹 그룹에서 '데이터 관리 가이드'를 발표하여 해당 내용에서 데이터 라이프 사이클 단계별 학습 모듈과 모범 사례를 제시하고 있고 제시된 학습 모듈이 RDM

또는 RDS를 위한 구성요소로서 서비스하고 있는 것을 확인하였다. <표 7>은 DataONE의 '데이터 관리 가이드'의 내용 중 데이터 라이프 사이클 단계별 하위 요소를 정리한 것이다.

DataONE 워킹그룹에 의해 발표된 '데이터 관리 가이드'는 데이터 라이프 사이클(Data Life Cycle) 단계를 중심으로 작성되었다. 첫 번째, '계획(Plan)' 단계는 데이터 라이프 사이클 동안 프로세스와 리소스를 매핑하는 단계로서 데이터 관리 계획을 세우고 데이터 정책 및 지속가능성 계획을 지원한다. 두 번째, '수집(Collect)' 단계는 참여자가 수집한 관측 데이터를 수집하고 구성하는 최적의 메커니즘을 결정하는 것으로 의사 결정 프로세스의 최종 결과는 데이터의 형식, 구성 및 처리 방법을 명확하게 설명하는 데이터 모델이다. 세 번째, '보장(Assure)' 단계는 교육 참여자, 정기적인 기기 교정 등을 통한 품질 보증 및 품질 관리 절차를 사용하여 데이터의 품질을 향상시키고 오류를 해결할 수 있는 잠재적 오류 및 기술을 식별하는 단계이다. 네

<표 7> DataONE의 데이터 라이프 사이클(DataONE Public Participation in Scientific Research Working Group 2013)

데이터 라이프 사이클 단계	세부 항목
계획(Plan)	<ul style="list-style-type: none"> • 데이터 관리 계획 • 데이터 관리 계획 도구
수집(Collect)	<ul style="list-style-type: none"> • 데이터 입력 및 처리(Manipulation)
보장(Assure)	<ul style="list-style-type: none"> • 데이터 품질 제어 및 보장
기술(Describe)	<ul style="list-style-type: none"> • 메타데이터 • 메타데이터 도구
보존(Preserve)	<ul style="list-style-type: none"> • 데이터 보호 및 백업
탐색(Discover)	<ul style="list-style-type: none"> • 데이터 공유
통합(Integrate)	<ul style="list-style-type: none"> • 분석 및 워크플로우
분석(Analysis)	<ul style="list-style-type: none"> • 분석 체크리스트 • 무료 분석 도구 예시

번째, ‘기술(Describe)’ 단계는 왜, 누가, 무엇을, 어디서, 어떻게 작성했는지를 설명함으로써 데이터를 문서화할 수 있으며, 메타데이터는 데이터 공유 및 재사용의 핵심으로 표준 및 소프트웨어와 같은 많은 도구를 사용하여 데이터를 설명할 수 있다. 다섯 번째, ‘보존(Preserve)’ 단계는 잠재적인 손실을 최소화하기 위해 단기간에 데이터를 보존하고 장기적으로 프로젝트 이해관계자와 다른 사람들이 향후 데이터에 액세스하고 해석하며 사용할 수 있도록 계획하는 단계로서 보존할 데이터, 보존할 위치 및 데이터와 함께 제공해야 하는 문서를 결정해야 한다. 여섯 번째, ‘탐색(Discover)’ 단계는 프로젝트 데이터에 가치를 추가할 수 있는 보완적인 데이터 세트를 식별하며, 데이터가 최대한 효과를 발휘할 수 있도록 도와주는 전략에는 리퍼지토리에 데이터를 등록하고 메타데이터 분석소에 데이터 설명을 추가하는 것이 포함된다. 일곱 번째, ‘통합(Integrate)’ 단계는 통합할 데이터 호환성 고려하고, 여러 데이터 셋 통합과 데이터 처리에서 이용된 단계를 문서화한다. 마지막으로, ‘분석(Analyze)’는 연구자, 참여자 및 기타 이해관계자를 위한 프로젝트 목표를 충족시키기 위해

데이터를 분석하며 데이터 탐색, 분석 및 시각화를 지원하는 많은 소프트웨어 도구를 사용할 수 있다.

3.5 요소 비교 분석

본 절에서는 기술한 내용을 토대로 DCC, ICPSR, ANDS 그리고 DataONE의 연구 데이터 관리 및 서비스를 비교 분석하였다. 분석 기준으로 활용한 것은 김지현(2014)이 대학도서관의 연구 데이터 관리 서비스를 분석하기 위해 사용했던 RDM 서비스의 9개 구성 요소를 활용하였다. <표 8>은 기준으로 활용한 연구 데이터 관리 서비스의 구성 요소와 정의를 나타낸 것이다.

RDM 서비스의 9개 요소는 ‘DMP 작성지원’, ‘데이터 파일 정리’, ‘데이터 기술’, ‘데이터 저장’, ‘데이터 공유 및 접근’, ‘데이터 보존’, ‘데이터 인용’, ‘데이터 관리 교육’ 그리고 ‘데이터 지적 재산권’으로 구성되어 있다.

<표 9>를 살펴보면, 9개의 RDM 서비스 구성 요소를 모두 포함하고 있는 기관은 ICPSR이다. 먼저, 연구 데이터 관리 및 서비스를 기준

<표 8> 연구 데이터 관리 서비스 구성 요소와 정의(김지현 2014, 174)

RDM 서비스 구성 요소	정의
DMP 작성 지원	연구자들의 DMP 작성을 보조, 지원하는 서비스
데이터 파일 정리	데이터 파일명 부여, 버전관리 및 파일포맷 선정 지원 서비스
데이터 기술	데이터에 적합한 메타데이터 표준 안내 서비스
데이터 저장	데이터의 저장 및 백업 관련 서비스
데이터 공유 및 접근	데이터의 공유 필요성과 공유방법에 대한 안내 서비스
데이터 보존	데이터 보존을 위한 리퍼지토리(Repository) 안내 서비스
데이터 인용	데이터 인용 방법과 DOI 안내 서비스
데이터 관리 교육	데이터 관리를 위한 개별 컨설팅, 워크숍 및 교육자료 제공 서비스
데이터 지적재산권	데이터와 관련된 저작권 및 라이선싱 안내 서비스

〈표 9〉 연구 데이터 관리 및 서비스 비교 분석

RDM 구성 요소 \ RDMS	DCC	ICPSR	ANDS	DataONE	비고
DMP 작성지원	○	○	○	○	■
데이터 파일 정리		○		○	
데이터 기술	○	○	○	○	■
데이터 저장	○	○	○	○	■
데이터 공유 및 접근	○	○	○	○	■
데이터 보존	○	○			
데이터 인용	○	○	○	○	■
데이터 관리 교육	○	○	○	○	■
데이터 지적 재산권		○	○	○	

으로 살펴보면, DCC는 ‘데이터 파일 정리’와 ‘데이터 지적 재산권’을 ANDS는 ‘데이터 파일 정리’, ‘데이터 보존’을 제외한 7개 구성 요소를 DataONE은 ‘데이터 보존’을 제외한 8개 구성 요소를 포함하고 있다.²⁾ RDM 구성 요소를 기준으로 연구 데이터 관리 및 서비스를 비교하면, 4개의 기관이 모두 포함하고 있는 구성 요소는 6개로서 ‘DMP 작성지원’, ‘데이터 기술’, ‘데이터 저장’, ‘데이터 공유 및 접근’, ‘데이터 인용’ 그리고 ‘데이터 관리 교육’이다. 4개의 기관에서 모두 다 포함하고 있는 요소만 기술하는 이유는 먼저, DCC는 RDM 서비스를 개발하기 위한 정책 관련 기관이므로 다른 어떤 기관보다 중요하다고 판단되어 해당 기관이 포함하지 않은 요소는 제외하였다. 다음, 특정 주제 분야의 RDM 서비스인 DCC와 ICPSR의 경우, 전 분야를 커버하는 데 한계가 있어 해당 기관만 포함하고 있는 요소도 기술 대상에서 제외시켰다.

다음의 4장에서는 6개의 구성 요소를 기준으

로 DCC, ICPSR, ANDS 그리고 DataONE에서 제시된 내용을 분석하고자 한다.

4. 요소별 기능적 요구사항 도출

이번 장에서는 3장에서 도출한 연구 데이터 관리 및 서비스의 공통 구성 요소인 ‘DMP 작성 지원’, ‘데이터 기술’, ‘데이터 저장’, ‘데이터 공유 및 접근’, ‘데이터 인용’ 그리고 ‘데이터 관리 교육’ 등 6개의 구성 요소에 대한 해당 내용을 기관별로 비교 분석하여 제시하였다.

4.1 DMP 작성 지원

데이터 관리 계획(Data Management Plan)은 어떤 데이터가 생성되며 데이터에 적용되는 정책은 무엇이며 데이터에 대한 소유권 및 액세스 권한, 데이터 관리 방법, 필요한 시설 및 장비 그리고 이러한 각 활동에 대한 책임은

2) DCC의 경우, RDM 서비스 개발방법에 대한 가이드가 2013년도에 발표된 후, 2015년에 데이터셋 인용방법에 대한 가이드를 발표하였다.

누구에게 있는지를 기술하는 문서이다(ANDS 2017). 여기에서는 DMP에 대한 각 기관별 하위 요소를 비교 분석하여 공통된 항목과 추가 항목을 도출하였다. <표 10>은 DCC, ICPSR, ANDS 그리고 DataONE 등에서 제시하는 데이터 관리 계획(Data Management Plan)의 하위 요소를 비교한 것이다.

분석 기준은 Science Europe(2018)의 「연구 데이터 관리의 국제적 협력」이라는 보고서에서 제시한 DMP를 위한 핵심 요구사항을 기준으로 정했다. 기준 항목에 따라 4개의 기관에서 제시하고 있는 항목을 표시하였으며 해당 기준 항목에 중복된 경우 해당 항목을 모두 표

시하였다. 제시한 기준의 핵심 요구사항에는 '현용 데이터의 데이터 설명 및 수집 또는 재사용', '도큐멘테이션 및 데이터 품질', '연구 과정 중 저장 및 백업', '법적 및 윤리적 요구 사항, 행동 강령', '데이터 공유 및 장기 보존' 그리고 '데이터 관리 책임 및 정보자원' 등 6개의 항목으로 구성되었다. 이 기준으로 DCC, ICPSR, ANDS 그리고 DataONE 등 4개의 기관을 분석한 결과 DMP를 위한 핵심 요구사항에 포함된 6개의 항목을 모두 포함하고 있는 것으로 나타났다. 특히 6개의 기준 항목 이외에 ICPSR은 '예산'을 ANDS는 '예산'과 '서지관리'를 DataONE은 '예산'을 추가적으로 제시하고 있다.

<표 10> DMP 구성요소 비교

DMP 구성요소 \ RDMS	DCC	ICPSR	ANDS	DataONE
현용 데이터의 데이터 설명 및 수집 또는 재사용	• 데이터 수집	• 현용 데이터 • 포맷 • 데이터 기술	• 현용 데이터 조사 • 생성될 데이터 • 파일 포맷	• 데이터 정보 & 데이터 포맷
도큐멘테이션 및 데이터 품질	• 도큐멘테이션과 메타데이터	• 메타데이터 • 품질 보장 • 데이터 정리 • 품질 보장	• 메타데이터 • 데이터 정리	• 메타데이터 콘텐츠 & 포맷
연구 과정 중 저장 및 백업	• 저장과 백업	• 저장 및 백업 • 보안	• 백업 • 접근 및 보안 • 저장	• 데이터 정보 & 데이터 포맷
법적 및 윤리적 요구 사항, 행동 강령	• 윤리 및 법규 준수	• 지적재산권 • 윤리 및 프라이버시 • 법적 요구사항	• 데이터 소유자 및 이해관계자	• 접근, 공유, 재사용 정책
데이터 공유 및 장기 보존	• 선택과 보존 • 데이터 공유	• 접근 및 공유 • 아카이빙 및 보존 • 이용자 • 선택 및 보유 기간	• 데이터 공유, 출판 및 아카이빙 • 파괴	• 접근, 공유, 재사용 정책 • 장기 저장 & 데이터 관리
데이터 관리 책임 및 정보자원	• 책임과 정보자원	• 책임	• 책임	• 데이터 정보 & 데이터 포맷 • 접근, 공유, 재사용 정책 • 장기 저장 & 데이터 관리
		• 예산	• 예산 • 서지관리	• 예산

4.2 데이터 기술

‘데이터 기술’ 요소는 3장의 연구 데이터 관리 및 서비스를 분석하기 위한 기준으로 사용된 9개 요소 중 하나로서 데이터에 적합한 메타데이터 표준 안내 서비스를 가리킨다. 이번 절에서는 분석한 4개의 서비스를 대상으로 ‘데이터 기술’의 내용을 비교하였다(〈표 11〉 참조).

4개의 기관에서 분석한 ‘데이터 기술’ 내용은 데이터에 적합한 메타데이터 표준 안내 서비스를 가리킨다. DCC는 메타데이터 정의와 메타데이터 생성 및 공개를 위한 시스템 설정을 ICPSR은 본래의 서비스 취지에 맞게 사회과학 데이터를 위한 메타데이터를 안내하고 있다. ANDS는 메타데이터 수집, 업데이트 및 유지 관리 그리고 데이터와 메타데이터의 연결을 위한 영구 식별자 사용 마지막으로 상호운용성 및 재사용을 위하여 하나 이상의 표준 메타데이터 사용을 제시하였다. 또한 DataONE은 펀더 등이 특정 메타데이터 사용을 권장할 경우 DMP에 정의하도록 제안하고 있고, 커뮤니티 내에서 혹은 일반적으로 사용하는 메타데이터의 사용 그리고 상호운용되는 메타데이터 표준을 사용하

도록 마지막으로 선호되는 표준이 없는 경우, 더블린 코어(Dublin Core) 등 표준 메타데이터를 사용하도록 제시하고 있다.

4.3 데이터 저장

분석기준으로 활용된 RDM 서비스 구성요소 중 ‘데이터 저장’은 데이터의 저장 및 백업 관련 서비스로서 〈표 12〉는 해당 내용에 대한 각 서비스별 내용을 비교한 것이다.

DCC 등 ‘데이터 저장 및 백업’ 관련 내용을 분석한 결과, DCC는 추가 스토리지를 제공하기 위해 연구 데이터 스토리지 할당 및 관리 절차 개발을 제안하고 다양한 범위의 연구 상황과 요구 사항을 충족하는 데이터의 생성, 관리 및 공유를 지원하는 유연한 RDM 시스템을 제공해야 한다고 제시한다. ICPSR은 해당 기관의 재난 계획에 포함된 백업에 대한 내용으로서 ICPSR이 보유한 모든 데이터에 대한 6개의 백업 사본 유지 관리와 정전시 백업 사이트를 기반으로 데이터 다운로드가 가능한 시스템을 제공하고 있다. ANDS는 프로젝트 초기에 데이터 저장 계획을 수립하도록 하고 있고 데이

〈표 11〉 데이터 기술 내용 비교

구성 요소	DCC	ICPSR	ANDS	DataONE
데이터 기술	<ul style="list-style-type: none"> • 필요한 메타데이터 정의 • 메타데이터를 위한 시스템 설정 • 메타데이터 생성 • 메타데이터 공개 	<ul style="list-style-type: none"> • 사회과학 데이터를 위한 메타데이터 안내 - DDI(Data Documentation Initiative) - 학습 및 변수 레벨 메타데이터(Study- and variable-level metadata) - 중요한 메타데이터 요소 	<ul style="list-style-type: none"> • 풍부하게 기술된 메타데이터 제시 • 메타데이터 수집, 업데이트 및 유지 관리 • 데이터와 메타데이터의 영구 연결을 지원하는 영구 식별자 • 하나 이상의 메타데이터 표준 사용 	<ul style="list-style-type: none"> • 관련 메타데이터 표준 식별 및 사용 - 특정 메타데이터 표준이 필요한 경우, DMP에 정의 - 커뮤니티 또는 범용 메타데이터 표준 사용 - 상호운용되는 메타데이터 표준 사용 - 선호되는 표준이 없는 경우, DC 표준 사용

〈표 12〉 데이터 저장 내용 비교

구성 요소	DCC	ICPSR	ANDS	DataONE
데이터 저장	<ul style="list-style-type: none"> • 현재 인프라 및 시스템이 충분한지 확인 • 추가 스토리지 제공하기 위한 사례 작성 • 스토리지 할당 및 관리 절차 개발 • 유연한 RDM 시스템 제공 	<ul style="list-style-type: none"> • 모든 데이터에 대한 6개의 백업 사본 유지 관리 • 클라우드 기반 서비스에 저장된 데이터 전송 시스템의 백업 사본 사이트 보유 • 정전시 백업 사이트에서 데이터 다운로드 가능 	<ul style="list-style-type: none"> • 프로젝트 초기에 데이터 저장 계획 수립 • 스토리지 솔루션 비교 기준 <ul style="list-style-type: none"> - 데이터의 가치와 재사용 가능성 - 발견 가능성, 큐레이션 및 안정적이고 규모가 크고 지속 가능한지 여부 	<ul style="list-style-type: none"> • 데이터 보호와 백업 관련 모범 사례 제시

터의 가치와 재사용 가능성 그리고 스토리지가 발견 가능성, 큐레이션 및 안정적이고 규모가 크고 지속 가능한지 여부를 확인하여 옵션에 맞는 유형을 선택하도록 제안하였다. 마지막으로, DataONE은 백업 관련한 모범 사례를 통해 정기적인 데이터 백업, 데이터 수집 프로세스에 맞는 최적화된 백업 전략 수립, 최소 2개 이상의 사본 생성 및 신뢰할 만한 위치 저장, 신뢰할 만한 장치 사용, 원본과 사본의 일치 여부 확인, 백업 사본으로부터 성공적인 복구를 보장하는 모든 절차 문서화 등을 제시하였다.

4.4 데이터 공유 및 접근

3장에서 언급된 연구 데이터 관리 및 서비스 분석 기준인 ‘데이터 공유 및 접근’은 데이터의 공유 필요성과 공유방법에 대한 안내 서비스로 정의된다. 〈표 13〉은 ‘데이터 공유 및 접근’에 대한 각 서비스별 내용을 비교한 것이다.

DCC는 해당 내용에 대해 데이터 선택 기준 수립, 기탁 가이드 개발, 데이터의 특성에 따라 다양한 수준의 큐레이션을 적용하기 위한 기준을 수립하도록 제안한다. ICPSR은 데이터 수

〈표 13〉 데이터 공유 및 접근 비교

구성 요소	DCC	ICPSR	ANDS	DataONE
데이터 공유 및 접근	<ul style="list-style-type: none"> • 보존할 데이터 선택 이유 • 연구자가 데이터를 선택하고 기탁하도록 권장 • 높은 수준의 가이드 • Advocacy 및 전공별 가이드 • 기탁 도구 • 기탁 동의 • 선택 결정을 위한 기준 수립 	<ul style="list-style-type: none"> • 데이터 수집 <ul style="list-style-type: none"> - 정량적 데이터 - 정성적 데이터 - 기타 데이터 유형 • 공유를 위한 데이터 준비 <ul style="list-style-type: none"> - 참여자 기밀유지 - 기밀유지 사례 - 이용 제한 데이터 수집 - 데이터 보호 구역 • 데이터 기탁 <ul style="list-style-type: none"> - 파일 포맷 - 현용 또는 2차 데이터 분석을 통한 파일 아카이빙 	<ul style="list-style-type: none"> • 데이터 공유 고려 사항 <ul style="list-style-type: none"> - 정보 시트 및 동의서는 데이터 공유 가능성 제시 - 민감한 데이터는 중재/통제된 접근 조건을 통해 공유되어 접근 규제 - 재사용을 위한 데이터의 경우, 라이선스 포함 	<ul style="list-style-type: none"> • 데이터 생성 시 고려사항 <ul style="list-style-type: none"> - 프라이버시 및 기밀 유지 - 저작권 및 지적재산권 - 라이선스

집의 종류를 정량적, 정성적 그리고 기타 데이터 유형으로 구분하여 데이터 수집 전략을 제공하고 연구자가 데이터를 보존 및 배포하기 위해 준비해야 고려사항으로 연구 참여자 기밀유지, 기밀유지 사례를 그리고 가장 중요한 기밀 데이터에 대한 액세스 권한을 부여하기 위하여 데이터 보호 구역 설정을 제안했다. 또한, 데이터 세트를 보존하려면 다른 사람들이 읽을 수 있는 방식으로 구성되어야 하며 원본 데이터 수집을 포함하지 않거나 하나 이상의 기존 소스의 데이터를 결합하는 프로젝트의 경우 기존 데이터 생산자의 허가를 받아 제출할 것을 제시하고 있다. ANDS는 데이터 공유를 위해 연구 참여자의 동의서에 데이터 공유 가능성을 포함해야 하며 민감한 데이터의 경우 중재 / 통제된 액세스 조건을 통해 공유되어 접근을 규제할 수 있으며 재사용을 위한 데이터에는 재사용 조건을 나타내는 라이선스가 있어야 한다고 제안한다. 마지막으로 DataONE은 공유를 위한 데이터 생성시 필요한 사항으로 기관 정책 내에 프라이버시 및

기밀유지에 대한 내용을 포함하도록 하고 있으며 저작권 및 지적재산권의 확인 그리고 데이터에 라이선스를 부여하여 다른 사람들이 이용할 때 조건을 명확히 하도록 제시하고 있다.

4.5 데이터 인용

‘데이터 인용’은 데이터 인용 방법과 DOI 안내 서비스로서 4개의 연구 데이터 관리 및 서비스에서 모두 포함하고 있다. 여기에서는 각 서비스별 데이터 인용에 포함되는 요소를 중심으로 정리하였다.

〈표 14〉에서 나타낸 내용을 살펴보면, DCC가 데이터 인용을 위한 구성요소가 가장 많음을 알 수 있다. DCC는 해당 출처를 통해 4개의 논문에서 언급한 구성 요소를 모두 나타낸 것이다. 이 중 인용문에 포함되어야 하는 요소로 저자, 타이틀 및 날짜, 위치 및 출판사를 제시하고 있다 (Ball and Duke 2015). 4가지 서비스에 공통적으로 확인되는 구성 요소는 저자, 발행일, 타이틀,

〈표 14〉 데이터 인용 내용 비교

구성 요소	DCC	ICPSR	ANDS	DataONE
데이터 인용	<ul style="list-style-type: none"> • 데이터 인용 구성요소 <ul style="list-style-type: none"> - 저자 - 출판일 - 타이틀 - 에디션(Edition) - 버전 - 기능 이름(Feature name) 및 URI - 정보자원 유형(database, dataset) - 출판사 - Unique numeric fingerprint (UNF) - 식별자 - 위치(영구적 URL) • 데이터 인용 원칙 	<ul style="list-style-type: none"> • 데이터 인용 구성요소 <ul style="list-style-type: none"> - 저자 - 타이틀 - 배포자 - 일자 - 버전 - 영구 식별자(DOI, URN, 핸들시스템) 	<ul style="list-style-type: none"> • 데이터 인용 구성요소 <ul style="list-style-type: none"> - 저자 - 발행년도 - 타이틀 - 버전 - 출판사 - DOI • 데이터 인용 지원 <ul style="list-style-type: none"> - 연구 평가 프로세스에 데이터 포함 - DCI(데이터 인용 색인) - RDA를 통한 국제 이니셔티브에 기여 - DOI 서비스 제공 	<ul style="list-style-type: none"> • 데이터 인용 구성요소 <ul style="list-style-type: none"> - 모든 참여자 - 데이터셋 발행 날짜 - 데이터셋 타이틀 - 매체 또는 URL - 데이터 출판사 - 식별자(DOI)

버전, 출판사, DOI 등 6개이다. 데이터 인용을 위한 구성 요소 이외에 DCC는 데이터 인용에 대한 8가지 원칙을 제시하고 있다. 이 원칙은 연구 데이터 연합(Research Data Alliance) 등 다양한 프로젝트의 대표가 포함된 FORCE 11 데이터 통합 그룹(Data Citation Synthesis Group)에서 발표한 것으로 중요성, 크레딧 및 기여, 근거, 고유 식별, 접근, 지속, 특이성 및 검증 가능성 그리고 상호 운용성 및 유연성 등을 포함한다. ANDS는 연구자들의 데이터 인용을 지원하기 위해 데이터 출판물을 주요 연구 성과물로 인식하여 연구 평가에 데이터를 포함하였으며 데이터 인용 색인(Data Citation Index)를 이용하여 데이터 인용을 추적한다. 또한, 데이터셋 및 컬렉션에 DOI를 할당하는 서비스를 제공하고 있다(ANDS Data Citation 2019).

4.6 데이터 관리 교육

DCC, ICPSR, ANDS 그리고 DataONE 등 연구 데이터 관리 및 서비스의 공통된 마지막 구성 요소는 '데이터 관리 교육'이다. '데이터 관리 교육'은 데이터관리를 위한 개별 컨설팅, 워크숍 및 교육자료 제공 서비스에 대한 내용이다.

관련된 내용은 <표 15>에서 기술하였다.

DCC는 RDM 가이드 웹사이트를 제공하기 위해 기존의 기관 지원에 대한 세부정보를 수집하고 안내 데스크 등을 통한 지원 제공, 연구자 및 지원 인력을 재교육하기 위한 RDM 교육 제공 그리고 심층적인 컨설팅 서비스를 제공한다. 또한 교육 대상을 두 가지 범주로 분류하여 특정 학문 분야 또는 대학원생을 대상으로 하는 연구자를 위한 과정과 사서와 같은 지원 인력을 재교육하기 위한 지속적인 전문 개발 과정(Continuing Professional Development)을 제시한다. 사회과학 데이터 리파지토리인 ICPSR은 실제 데이터를 이용한 오프라인 교육을 진행하며 DCC와 유사하게 교육 대상 범주를 학생과 교사로 나누어 교육 프로그램을 제시하고 있다. 학생을 위한 학습자료는 저널 기사를 읽는 법, SPSS 산출물 해석 가이드, 데이터 인용 그리고 데이터 관리 계획 등을 포함하며, 교사를 위한 학습자료는 다양한 주제의 강의 동영상과 학부생을 위한 외부 정보자원을 제공한다. ANDS는 연구 데이터를 위한 아웃리치 프로그램과 학습 자료를 포함하고 있는 연구 데이터 리파지토리 목록을 제시하며 특정 영역 및 분야에 맞춘 도메인별 교육 등을 제시한다. DataONE

<표 15> 데이터 관리 교육 내용 비교

구성 요소	DCC	ICPSR	ANDS	DataONE
데이터 관리 교육	<ul style="list-style-type: none"> • 가이드와 헬프데스크 • 이용자별 교육 • 컨설팅 서비스 • 데이터 관리 교육 범주 <ul style="list-style-type: none"> - 연구자를 위한 과정 - 지원 인력을 재교육하기 위한 지속적인 전문 개발 	<ul style="list-style-type: none"> • 오프라인 교육 • 학생을 위한 학습자원 • 교사를 위한 학습자원 	<ul style="list-style-type: none"> • 연구 데이터를 위한 아웃리치 프로그램 • 학습 자료를 포함하고 있는 연구 데이터 리파지토리 목록 • 스킬(Skill) 	<ul style="list-style-type: none"> • 온라인 교육 <ul style="list-style-type: none"> - 데이터 관리 모듈 - 웹 세미나 - 스크린캐스트 튜토리얼 - 데이터 스토리 - 교육평가 • 교육 활동 <ul style="list-style-type: none"> - In-Person

은 온라인 및 오프라인을 통해 데이터 관리를 위한 모범 사례, 교육 자료 제공, 도구 및 정보 자원에 대한 링크 등 데이터 책임에 대한 커뮤니티 교육을 제공하고 있다.

5. RDMS 모델 제안

이번 장에서는 연구 데이터 관리 및 서비스를 위한 모델을 제시하기 위하여 DCC, ICPSR, ANDS 그리고 DataONE의 연구 데이터 관리 및 서비스의 공통된 6개의 구성 요소와 해당 내용을 제시했다. 구성 요소는 ‘DMP 작성 지원’, ‘데이터 기술’, ‘데이터 저장’, ‘데이터 공유 및 접근’, ‘데이터 인용’ 그리고 ‘데이터 관리 교육’이며 해당 내용은 각 기관이 제시하는 내용을 종합하여 <표 16>에서 기술하였다. DMP 작성

의 경우, 전 세계적으로 펀딩 기관에 DMP 제출이 의무화되어가고 있으며 국가 또는 기관별로 DMP의 세부항목이 표준화되고 있어 Science Europe(2018)이 제시한 핵심 요구사항을 기준으로 4개 기관의 DMP 세부 항목을 비교 및 분석하였다. DMP 작성 지원의 세부 항목과 내용은 해당 분석 기준을 활용된 Science Europe(2018)의 핵심 요구사항에서 제시한 세부 항목과 내용을 참조하였다. 반면, 나머지 ‘데이터 기술’ 등 5개의 구성 요소는 분석된 4개의 기관별로 조금씩 차이가 있으며 DMP처럼 공통된 항목과 추가 항목을 비교 및 분석하는 것이 적절치 않다고 판단되어 각 기관에서 제시하고 있는 세부 항목과 내용을 핵심요소 및 기능적 요건으로 종합 기술하였다. <표 16>은 연구 데이터 관리 및 서비스를 위한 핵심요소 및 기능적 요건을 제시한 것이다.

<표 16> 연구 데이터 관리 및 서비스를 위한 핵심요소 및 기능적 요건

구성 요소	세부 항목	내용	비고
DMP 작성 지원	• 현용 데이터의 데이터 설명 및 수집 또는 재사용	• 새로운 데이터를 수집하거나 생산하는 방법, 기존 데이터의 재사용 • 수집되거나 생산된 데이터의 종류, 포맷, 볼륨 등	필수
	• 도큐멘테이션 및 데이터 품질	• 메타데이터 및 도큐멘테이션(데이터 수집 및 정리 방법) • 데이터 품질 관리 수단	"
	• 연구 과정 중 저장 및 백업	• 데이터와 메타데이터의 저장 및 백업 • 중요한 데이터에 대한 데이터 보안 및 보호	"
	• 법적 및 윤리적 요구 사항, 행동 강령	• 개인 데이터 및 보안에 관한 법률 준수 • 지적재산권 및 소유권 등 기타 법적 문제 • 윤리적 문제와 윤리 강령	"
	• 데이터 공유 및 장기 보존	• 데이터 공유 방법 및 시기, 공유 및 엠바고에 대한 제한 • 보존을 위한 데이터 선택 및 데이터의 장기간 보존(리퍼지토리, 아카이브) • 소프트웨어 도구 등 데이터 접근 방법 • DOI 적용 방법	"
	• 데이터 관리 책임 및 정보자원	• 데이터 관리에 대한 책임 • FAIR 데이터(검색, 접근, 상호운용 그리고 재사용 가능)	"
	• 예산	• 아카이빙, 데이터 큐레이션, 메타데이터 생성 등을 위한 데이터와 도큐멘테이션 준비 비용(하드웨어와 소프트웨어) • 인력 및 연구 보조 시간	옵션
	• 서지 관리	• 서지 관리 도구 및 레퍼런스 공유 방법	"

구성 요소	세부 항목	내용	비고
데이터 파일 정리	<ul style="list-style-type: none"> • 메타데이터 정의 • 메타데이터 시스템 • 메타데이터 표준 • DOI 	<ul style="list-style-type: none"> • 풍부하게 기술된 메타데이터 제시 • 메타데이터 생성 및 공개를 위한 시스템 설정 • 메타데이터 수집, 업데이트 및 유지 관리 • 관련 메타데이터 표준 확인 및 사용 • 상호운용성을 위한 메타데이터 표준 사용 • 특정 메타데이터 표준이 필요한 경우, DMP에 정의 • 커뮤니티 또는 범용 메타데이터 표준 사용 • 데이터와 메타데이터의 영구 연결을 지원하는 식별자 	
데이터 저장	<ul style="list-style-type: none"> • 데이터 저장 계획 수립 • 스토리지(storage) 할당 및 관리 • 스토리지 솔루션 비교 • 데이터 보호 및 백업 정책 	<ul style="list-style-type: none"> • 프로젝트 초기에 데이터 저장 계획 수립 • 필요한 경우, 추가 스토리지 제공하기 위한 인프라 및 시스템 확인 • 스토리지 할당 및 관리 절차 개발 • 스토리지 솔루션 비교 기준 <ul style="list-style-type: none"> - 데이터의 가치와 재사용 가능성 - 탐색 가능성, 쿼레이션 및 안정적이고 규모가 크고 지속 가능한지 여부 • 모든 데이터에 대한 백업 사본 유지 관리 • 클라우드 기반 서비스에 백업 사본 사이트 보유 • 정전시 백업 사이트에서 데이터 다운로드 서비스 	
데이터 공유 및 접근	<ul style="list-style-type: none"> • 데이터 수집 및 선택 • 데이터 공유를 위한 준비 • 데이터 기탁 • 데이터 공유 고려사항 	<ul style="list-style-type: none"> • 수집할 데이터 유형 결정 • 선택 결정을 위한 기준 수립 • 공유를 위한 데이터 준비 <ul style="list-style-type: none"> - 프라이버시 및 기밀 유지 - 저작권 및 지적재산권 확인 - 이용제한 데이터 수집 - 데이터 보호 구역(Enclave) • 데이터 기탁 <ul style="list-style-type: none"> - 기탁을 위한 파일 포맷 - 기탁 동의서 작성 및 기탁 도구 제시 • 데이터 공유 고려사항 <ul style="list-style-type: none"> - 동의서 작성 시 데이터 공유 가능성 제시 - 민감한 데이터는 중재/통제된 접근 조건을 통해 공유되어 접근 규제 - 재사용을 위한 데이터의 경우, 라이선스 포함 	
데이터 인용	<ul style="list-style-type: none"> • 데이터 인용 구성 요소 • 데이터 인용 지원 	<ul style="list-style-type: none"> • 데이터 인용 구성 요소 <ul style="list-style-type: none"> - 저자 - 발행일 - 타이틀 - 버전 - 출판사 - DOI • 데이터 인용 지원 <ul style="list-style-type: none"> - 연구 평가 프로세스에 데이터 포함 - 데이터 인용 색인(Data Citation Index) - RDA를 통한 국제 이니셔티브에 기여 - DOI 서비스 제공 	
데이터 관리 교육	<ul style="list-style-type: none"> • 가이드와 헬프 데스크 • 컨설팅 서비스 • 데이터 관리 교육 범주(이용자별 교육) • 온라인 및 오프라인 교육 	<ul style="list-style-type: none"> • 연구 데이터 관리 및 서비스 안내 웹 사이트 제공 • 헬프 데스크를 통한 지원 제공 조정 • 이용자별 교육 <ul style="list-style-type: none"> - 연구자를 위한 과정 - 지원 인력을 재교육하기 위한 지속적인 전문 개발 • 온라인 교육 <ul style="list-style-type: none"> - 데이터 관리 모듈 - 웹 세미나(Webinar) - 스크린캐스트 튜토리얼 - 데이터 스토리 • 오프라인 교육 <ul style="list-style-type: none"> - 데이터를 활용한 데이터 관리 교육 - 특정 커뮤니티를 위한 집합 교육 	

첫 번째, 'DMP 작성 지원'이다. 이 데이터 관리 계획의 해당 세부 항목은 4장의 1절에서 분석의 기준인 Science Europe(2018)이 제시한 6개 세부 항목과 ICPSR, ANDS 그리고 DataONE 등에서 추가로 제시한 2가지 항목 등 8개의 세부 항목으로 구성되었다. 8개의 세부 항목 중 6개는 '필수' 항목으로 DMP 작성시 반드시 포함해야 할 내용이며 그리고 2개는 '옵션' 항목으로 DMP 작성시 작성자가 선택할 수 있도록 하였다. '필수' 항목은 '현용 데이터의 데이터 설명 및 수집 또는 재사용', '도큐멘테이션 및 데이터 품질', '연구 과정 중 저장 및 백업', '법적 및 윤리적 요구 사항, 행동 강령', '데이터 공유 및 장기 보존' 그리고 '데이터 관리 책임 및 정보 자원'이며 '옵션' 항목은 '예산'과 '서지 관리'이다. 필수 항목은 분석 대상인 4개의 기관에서 모두 포함하고 있는 항목이며 옵션 항목은 1개 이상의 기관에서 포함하고 있는 항목이다. 각 세부 항목별 내용은 <표 16>에 기술되어 있다.

두 번째, '데이터 파일 정리' 항목은 데이터에 적합한 메타데이터 표준 안내 서비스로서 풍부하게 기술된 메타데이터 제시, 메타데이터 생성 및 공개를 위한 시스템 설정, 메타데이터 수집과 업데이트 및 유지 관리, 상호운용성을 위한 메타데이터 표준, 데이터와 메타데이터의 영구적 연결을 지원하는 식별자(DOI) 등을 세부 내용을 제시하였다.

세 번째, '데이터 저장'은 데이터 저장 및 백업 관련 서비스로서 '데이터 저장 계획 수립', '스토리지(Storage) 할당 및 관리', '스토리지 솔루션 비교' 그리고 '데이터 보호 및 백업 정책' 등으로 세부 항목을 구성하였다. 프로젝트 초기에 데이터 저장 계획을 수립하도록 제시하

며, 추가 스토리지가 필요한 경우 스토리지 할당 및 관리 절차를 개발하도록 제안하였다. 또한 '데이터 보호 및 백업 정책'으로는 모든 데이터에 대한 백업 사본 유지 관리와 정전과 같은 비상시 백업 사이트에서 데이터를 다운로드할 수 있는 서비스를 포함하였다.

네 번째, '데이터 공유 및 접근'은 '데이터 수집 및 선택', '데이터 공유', '데이터 기탁', '데이터 공유시 고려사항'에 대한 것이다. 여기에는 데이터 유형 결정, 데이터 선택 결정을 위한 기준 수립, 데이터 공유를 위한 데이터 준비, 데이터 기탁을 위한 파일 포맷, 기탁 동의서 작성 및 기탁 도구 제시 그리고 데이터 공유시 고려사항으로 민감한 데이터는 중재/통제된 접근 조건을 통해 공유되어 접근을 규제하고 재사용을 위한 데이터의 경우, 라이선스를 포함해야 한다는 고려사항을 포함하고 있다.

다섯 번째, '데이터 인용'에 대한 것으로 '데이터 인용 구성 요소'와 '데이터 인용 지원' 등 2개의 세부 항목으로 구성되었다. 해당 내용에서는 '데이터 인용 구성 요소'로 저자, 발행일, 타이틀, 버전, 출판사 그리고 DOI를 반드시 포함하도록 제안하였고, '데이터 인용 지원'은 실제 기관에서 연구 평가 프로세스에 논문 외에 데이터를 포함하도록 하였고 데이터 인용 색인(Data Citation Index)을 이용하여 인용 추적을 그리고 데이터셋 등에 DOI를 할당하는 서비스를 제시하였다.

마지막으로 '데이터 관리 교육'에서는 '가이드와 헬프 데스크', '컨설팅 서비스', '이용자별 교육' 그리고 온라인 및 오프라인 교육 내용을 세부 항목으로 제시하였다. 해당 내용으로 연구 데이터 관리 및 서비스 안내 웹 사이트 제공,

헬프 데스크를 통한 지원 제공, 연구자와 지원 인력(사서 등)을 구분하여 연구자를 위한 과정과 지원 인력의 재교육 과정 등으로 이용자별 교육을 제시하였다. 또한 온라인 및 오프라인 교육을 통하여 데이터 관리 모듈 및 성공적인 데이터 관리 경험 그리고 데이터를 활용한 데이터 관리 교육, 특정 커뮤니티를 위한 집합 교육을 제안하였다.

6. 결 론

데이터의 가치 증대, 연구 방법의 패러다임 변화 그리고 오픈 사이언스 등의 구체적 발현은 연구가 더 이상 기존의 학술지와 같은 텍스트 중심이 아닌 데이터 기반으로 전환되고 있음을 나타내고 있다. 특히 연구 데이터는 국가 과학기술 경쟁력을 좌우하는 원천이며, 궁극에는 4차 산업 혁명에서 주요 변화 동인으로 다뤄지는 만큼 국가에 의한 적극적인 관리가 시급한 실정이다. 선진국과 다르게 우리는 연구 데이터에 그 동안 미흡했던 게 사실이지만 현재는 법제도에 연구 데이터 정의와 데이터 관리 계획(DMP) 도입 및 DMP 정보 등록 근거가 신설되었다.³⁾ 또한 정부는 「국가 연구 데이터 공유·활용 전략안」을 통해 연구 데이터에 대한 관리와 활용에 대한 국가 정책 및 의지를 표명하였으며, 2018년 12월에 과학기술정보통신부 산하기관인 정보통신기술진흥센터와 국가과학기술연구회에서는 DMP 적용을 위한 이행방안을 마련하였다.

본 연구의 목적은 연구 데이터 관리 및 서비

스를 기획 및 운영하고자 할 때 해당 서비스를 위한 핵심요소의 기능적 요건을 도출하는 것이다. 이러한 목적을 달성하고자 아직 국내 연구가 미흡한 연구 데이터 관리 및 서비스를 위한 핵심요소와 해당 기능적 요건을 도출하기 위하여 DCC, ICPSR, ANDS 그리고 DataONE에 대한 서비스를 분석하였다. 분석 결과 도출된 핵심 요소는 'DMP 작성지원', '데이터 기술', '데이터 저장', '데이터 공유 및 접근', '데이터 인용' 그리고 '데이터 관리 교육' 등 6가지가 도출되었다. 도출된 6가지는 '연구 데이터 관리 및 서비스'에 대한 기획 및 운영하고자 할 때 우선순위로 고려될 수 있는 서비스로 이 6가지의 핵심 요소에 대한 기능적 요건은 5장에서 제시하였다.

그 동안 연구 데이터 관리 및 서비스 관련 선행연구는 해당 서비스의 구성 요소를 제시하거나 서비스 우선순위를 제시한 논문이 대부분이다. 본 연구에서 도출된 핵심 요소에 기능적 요건을 제시함으로써 실제 연구 데이터 관리 및 서비스를 구축 및 운영하고자 할 때 본 연구의 내용을 적용할 수 있을 것이다. 제시된 모델이 일반적인 연구 데이터 관리 및 서비스에 적용할 수 있으나 특정 기관이나 이용자 커뮤니티에 서비스를 적용하는 것은 한계가 있다. 추후 학문 분야별 또는 특정 이용자 커뮤니티 등으로 제한한 연구가 진행될 필요가 있다. 또한, 효과적인 연구 데이터 관리 및 서비스가 구축되기 위해서는 해당 도메인에서 사용되는 연구 데이터에 대한 분석이 이루어져야 하며 또한 해당 도메인과 관련된 커뮤니티의 다양한 이해관계자들의 의견이 적극적으로 반영이 되어야 할 것이다.

3) 국가연구개발사업의 관리 등에 관한 규정이 2019년 3월 19일에 개정되었음.

참 고 문 헌

- [1] 김선태. 2015. 연구패러다임 변화와 데이터 집중형 과학. 국립해양생물자원관, 2015년 11월 26일. [online] [cited 2019. 7. 2.] <<https://www.slideshare.net/suntaekim/ss-55398568>>
- [2] 김선태, 이태영, 김용. 2012. 해양관측데이터 관리를 위한 결정요소 도출. 『정보관리연구』, 43(3): 97-115.
- [3] 김지현. 2012. 대학 내 연구자들의 연구데이터 관리에 관한 연구. 『한국도서관·정보학회지』, 43(3): 433-455.
- [4] 김지현. 2014. 대학도서관의 연구데이터관리서비스에 관한 연구. 『한국비블리아학회지』, 25(3): 165-189.
- [5] 박미영, 안인자, 김준모. 2018. 생명공학분야의 연구데이터 공유 사례에 관한 연구. 『한국비블리아학회지』, 29(1): 393-416.
- [6] 신은정 외. 2017. 『오픈 사이언스 정책의 도입 및 추진 방안』. 세종: 과학기술정책연구원, 정책연구 2017-08.
- [7] 심원식. 2016. 미국 대학도서관의 연구데이터 지원 서비스 사례 연구. 『한국문헌정보학회지』, 50(4): 311-332.
- [8] 안형준. 2017. 오픈사이언스에 대한 연구자 인식조사 연구. 『2017 기술경영경제학회 하계학술발표회 논문집』, 2017년 6월 23일~24일. 제주: 제주 KAL 호텔.
- [9] 예상준, 장호, 김선태. 2019. 한의학 연구 데이터 관리 및 공유를 위한 메타데이터 요소 설계. 『한국문헌정보학회지』, 53(2): 223-246.
- [10] 이상환. 2019. DMP(Data Management Plan), 연구데이터 공유와 활용의 시작. 『'19년 한국전문도서관협의회-국가정책정보협의회 공동 학술세미나』, 2019년 6월 27일. 여수: 디오션리조트 호텔: 31-46.
- [11] 최명석, 이승복, 이상환. 2017. 국내 과학기술분야 연구기관의 과학데이터 관리 현황. 『한국콘텐츠학회논문지』, 17(12): 117-126.
- [12] About DataONE. [online] [cited 2019. 7. 7.] <<https://www.dataone.org/about>>
- [13] About the DCC (Digital Curation Centre). [online] [cited 2019. 7. 6.] <<http://www.dcc.ac.uk/about-us>>
- [13] ANDS. 2017. *Data management plans*. [online] [cited 2019. 7. 13.] <https://www.ands.org.au/_data/assets/pdf_file/0011/690878/Data-management-plans.pdf>
- [14] ANDS About us. [online] [cited 2019. 7. 7.] <<https://www.ands.org.au/about-us>>
- [15] Ball, A. and Duke, M. 2015. 'How to Cite Datasets and Link to Publications'. DCC How-to

- Guides. Edinburgh: Digital Curation Centre.* [online] [cited 2019. 7. 7.]
〈<http://www.dcc.ac.uk/resources/how-guides>〉
- [16] Cox, A. M. et al. 2017. "Developments in research data management in academic libraries: Towards an understanding of research data service maturity." *Journal of the Association for Information Science and Technology*, 68(9): 2182-2200. [online] [cited 2019. 7. 8.]
〈<https://doi.org/10.1002/asi.23781>〉
- [17] Crouzier, T. 2015. *Science Ecosystem 2.0: How will Change Occur?*, European Commission. [online] [cited 2019. 7. 8.]
〈https://ec.europa.eu/research/innovation-union/pdf/expert-groups/rise/science_ecosystem_2.0-how_will_change_occur_crouzier_072015.pdf〉
- [18] Data citation. [online] [cited 2019. 7. 12.]
〈<https://www.andis.org.au/working-with-data/citation-and-identifiers/data-citation>〉
- [19] Data Management & Curation. [online] [cited 2019. 7. 7.]
〈<https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/index.html>〉
- [20] DataONE Public Participation in Scientific Research Working Group. 2013. *Data Management Guide for Public Participation in Scientific Research*. [online] [Cited 2019. 7. 8.]
〈<https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>〉
- [21] Fecher, B. and S. Friesike. 2013. "Open Science: One Term, Five Schools of Thought." *In: Opening Science* (edited by Bartling, S. and S. Friesike), New york: Springer. [online] [cited 2019. 7. 6.]
〈<https://pdfs.semanticscholar.org/1752/a86b1a6ba3e684f07115de52fb2c60d068a0.pdf>〉
- [22] Inter-university Consortium for Political and Social Research. [online] [cited 2019. 7. 6.]
〈https://en.wikipedia.org/wiki/Inter-university_Consortium_for_Political_and_Social_Research〉
- [23] 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.* [online] [cited 2019. 7. 8.] 〈<http://www.dcc.ac.uk/resources/how-guides>〉
- [24] Koltay, T. 2017. "Research 2.0 and research data services in academic and research libraries: Priority issues." *Library Management*, 38(6): 345-353. [online] [cited 2019. 7. 8.]
〈<https://www.emerald.com/insight/content/doi/10.1108/LM-11-2016-0082/full/html>〉
- [25] MERCURY PROJECT SOLUTIONS. 2013. *Research Data Management in Practice*. [online] [cited 2019. 7. 7]

- 〈https://www.andis.org.au/_data/assets/pdf_file/0009/394056/research-data-management-in-practice.pdf〉
- [26] OECD 2015. "Making Open Science a Reality." *OECD Science, Technology and Industry Policy Papers*. No. 25, Paris: OECD Publishing. [online] [cited 2019. 7. 8.] 〈<https://doi.org/10.1787/5jrs2f963zs1-en>〉
- [27] OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publishing. [online] [cited 2019. 7. 5.] 〈<http://www.oecd.org/sti/inno/38500813.pdf>〉
- [28] OECD. 2016. *OECD Science, Technology and Innovation Outlook 2016*. Paris: OECD Publishing. [online] [cited 2019. 7. 5.] 〈http://dx.doi.org/10.1787/sti_in_outlook-2016-en〉
- [29] Open Science. [online] [cited 2019. 7. 2.] 〈https://en.wikipedia.org/wiki/Open_science〉
- [30] 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. [online] [cited 2019. 7. 4.] 〈<http://dx.doi.org/10.21083/partnership.v12i2.4115>〉
- [31] Research Data Management: Why Manage Research Data?. [online] [cited 2019. 7. 9.] 〈https://libguides.ucd.ie/data/why_manage〉
- [32] SCIENCE EUROPE. 2018. *PRACTICAL GUIDE TO THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT*. [online] [cited 2019. 7. 8.] 〈<https://www.scienceeurope.org/our-resources/practical-guide-on-research-data-management/>〉
- [33] Steeleworthy, M. 2014. "Research data management and the canadian academic library: An organizational consideration of data management and data stewardship." Partnership: *The Canadian Journal of Library and Information Practice and Research*, 9(1): 1-11. [online] [cited 2019. 7. 4.] 〈<http://dx.doi.org.proxy.jbnu.ac.kr/10.21083/partnership.v9i1.2990>〉
- [34] What is Open Data?. [online] [cited 2019. 7. 8.] 〈<http://opendatahandbook.org/guide/en/what-is-open-data/>〉
- [35] Yu, H. H. (2017). "The role of academic libraries in research data service (RDS) provision." *The Electronic Library*, 35(4): 783-797. [online] [cited 2019. 7. 4.] 〈<https://www.emerald.com/insight/content/doi/10.1108/EL-10-2016-0233/full/html>〉

• 국문 참고자료의 영어 표기

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

- [1] Kim, Sun-Tae. 2015. Research Paradigm Shift and Data-Oriented Science. National Marine Biodiversity Institute of Korea, 26 November 2015. [online] [cited 2019. 7. 2.]
 <<https://www.slideshare.net/suntaekim/ss-55398568>>
- [2] Kim, Sun-Tae, Lee, Tae-Young and Kim, Yong. 2012. "Deriving the Determining Factor for the Management of Oceanographic Data." *Journal of Information Management*, 43(3): 97-115.
- [3] Kim, J. 2012. "A Study on University Researchers' Data Management Practices." *Journal of Korean Library and Information Science Society*, 43(3): 433-455.
- [4] Kim, J. 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.
- [5] Park, M., Ahn, I. and Kim, J. 2018. "A Study on Use Case of Research Data Sharing in Biotechnology." *Journal of the Korean BIBLIA Society for library and Information Science*, 29(1): 393-416.
- [6] Shin, Eun-jung et al. 2017. *Introduction and Promotion Plan of Open Science Policy*. Sejong: Science and Technology Policy Institute, Policy Research 2017-08.
- [7] 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.
- [8] Ahn, Hyung-joon. 2017. Researcher's Perception Survey on Open Science. *Proceedings of 2017 Summer Conference of Korea Technology Management Economics*, June 23-24, 2017. Jeju: Jeju KAL Hotel.
- [9] Yea, S. J., Jang, H. and Kim, S. 2019. "Metadata Element Design for Korean Medicine Research Data Management and Re-use." *Journal of the Korean Society for Library and Information Science*, 53(2): 223-246.
- [10] 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: 31-46.
- [11] Choi, M. S., Lee, S. B. and Lee, S. 2017. "Research Data Management of Science and Technology Research Institutes in Korea." *The Journal of the Korea Contents Association*, 17(12): 117-126.