

데이터 관리 계획의 국내 현황 및 과제

Current Status and Issues of Data Management Plan in Korea

최명석, 이상환
한국과학기술정보연구원 연구데이터공유센터

Myung-seok Choi(mschoi@kisti.re.kr), Sanghwan Lee(sanglee@kisti.re.kr)

요약

최근 디지털 기술의 발전으로 연구 패러다임이 데이터 중심으로 변화하고 있다. 연구데이터의 국가적 관리와 활용은 연구의 투명성과 효율성을 제고하고, 나아가 데이터 기반으로 급변하는 사회에 대비하기 위한 필수 요소이다. 주요 선진국에서는 공적 지원에 의해 수행된 연구에서 생산되는 연구데이터의 공유·활용을 위한 정책과 인프라 구축이 적극적으로 추진되고 있다. 국내에서도 2019년 국가연구개발과제 제안시 데이터 관리 계획을 제출하도록 관련 규정이 개정되었다. 데이터 관리 계획 제도가 효과적으로 실행되고 지속성을 유지하기 위해서는 연구자에게 다양한 지원 여건이 필요하다. 또한 데이터 관리 계획 가이드라인과 이행 절차는 국가 또는 기관 차원의 연구데이터 관리·활용을 위해 필수적이다. 이 연구에서는 데이터 관리 계획의 개요와 국내의 추진 현황 및 이슈를 살펴보고, 작성 가이드라인과 체크리스트, 적용프로세스, 그리고 연구기관에서의 이행 방안을 제시한다.

■ 중심어 : | 연구데이터관리 | 오픈사이언스 | 오픈연구데이터 | 데이터관리계획 | 데이터 리포지터리 |

Abstract

With the recent development of digital technology, the research paradigm is evolving towards data-driven. National management and utilization of research data is a key element not only to enhance research transparency and efficiency, but also to prepare for a data-driven society. Policies and infrastructure for sharing and utilization of research data from publicly-funded research are being actively promoted worldwide. In Korea, related regulations were recently revised to mandate to submit a data management plan (DMP) when proposing a national R&D project. In order to effectively implement the sustainable DMP system, researchers need various support. In addition, guidelines and implementation procedures are essential for management and utilization of research data at the national or institutional level. In this paper, we provide an overview of the data management plan, examine the current status and issues in Korea, and suggest a template and checklists of data management plan, and an implementation procedure at research institutes.

■ keyword : | Scientific Data Management | Research Data Management | Open Science | Open Research Data | Data Repository |

* 이 연구는 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행된 연구입니다.

접수일자 : 2020년 05월 13일
수정일자 : 2020년 06월 08일

심사완료일 : 2020년 06월 08일
교신저자 : 이상환, e-mail : sanglee@kisti.re.kr

I. 서론

디지털 기술의 발전으로 데이터 분석과 활용의 중요성이 증가함에 따라 데이터 활용 역량이 4차 산업시대에서 과학기술 경쟁력의 핵심으로 대두되고 있다. 이러한 디지털 기술은 연구 성과의 확산 방식 뿐만 아니라 연구의 추진 방식 자체를 변화시키고 있다. 이에 따라 연구개발 패러다임도 데이터 중심으로 변화하고 있다 [1].

이와 더불어 공적 지원에 의해 생산된 연구결과물에 대한 쉬운 접근과 활용을 위한 오픈 사이언스 패러다임이 오픈 연구데이터로 확대되고 있다[2-4]. 오픈 연구데이터는 연구 과정에서 생산된 연구데이터에 대한 자유로운 접근과 재사용을 허용하는 것이다[2]. 주요 선진국에서는 공공지금이 투입된 연구과제로부터 생산된 데이터의 체계적인 관리와 쉬운 접근, 재사용을 통한 가치 창출을 위해 데이터 관리 계획(Data Management Plan, DMP) 등 오픈 연구데이터 정책을 시행하고 있다[5-8].

미국에서는 연간 1억 달러 이상의 연방정부 R&D 예산을 지출하는 모든 연방 기관은 공공 연구결과물에 대한 공공접근계획(Public Access Plan) 제출을 의무화하였다. 특히, 국립보건원(NIH)는 2003년 미화 50만 달러를 초과하는 연구과제의 경우 데이터공유계획(Data Sharing Plan) 제출을 의무화하고 있으며, 국립과학재단(NSF)도 2011년 1월부터 모든 연구과제에 대해 DMP를 첨부하도록 의무화하였다.

영국은 영국연구혁신기구(UKRI)의 '데이터 정책에 관한 공동 원칙(Common Principles on Data Policy)'과 '오픈 연구데이터에 대한 협약(Concordat on Open Research Data)'을 기반으로 7개의 연구위원회가 독립적으로 분야별 데이터 관리·공유 정책, 규제 규정 등을 제시하고 있다. 6개의 연구위원회에서 DMP 의무제출 규정을 시행중이며, 나머지 1개인 공학및물리학연구위원회(EPSRC)는 출판물에 연구데이터 정보와 접근방안 기재를 의무화하고 있다.

호주의 경우 2007년 '책임 있는 연구수행을 위한 국가 지침(Australian Code for the Responsible Conduct of Research, the Code)'을 발표하여 연구

산출물의 생산·관리 규정을 제시하고 있다. 호주연구위원회(ARC)는 2014년부터 연구비를 지원 받는 기관에 대하여 DMP 제출을 의무화하고 있으며, DMP 세부 지침과 템플릿 제공의 책임은 대학 등 개별 연구기관에 부여하고 있다.

국내는 연구데이터 관련 제도, 인력, 인프라 등이 전반적으로 미비한 상황이다. 공공데이터와 일부 연구데이터에 대한 공유·활용 제도가 시행되고 있으나, 국가연구개발과제의 결과물로서 국가 차원에서 연구데이터의 공유·활용을 위한 법적 기반은 부재한 상황이다[8]. 또한 국가가 지정한 특정 데이터센터를 중심으로 일부 연구데이터를 수집·관리하고 있으나, 데이터의 활용에는 한계가 있다. 개별 연구기관이 소속 연구자들이 연구데이터를 원활하게 공유·활용할 수 있도록 지원할 수 있는 역량도 아직 부족하다.

연구데이터의 국가적 관리와 활용은 연구의 투명성과 효율성을 제고하고, 나아가 데이터 기반으로 급변하는 사회에 대비하기 위한 필수 요소이다. 최근, 우리나라도 이런 연구데이터의 중요성을 인식하여 국가 정책 의제로 발표하고, 연구 성과물로서의 연구데이터의 관리와 활용을 위한 법적 근거를 마련하였다. 2019년 정부는 「국가연구개발사업의 관리 등에 관한 규정」(이하 '공동관리규정')에 연구데이터와 DMP의 정의, DMP 제출 의무사항 등을 반영하였다.

DMP 제도가 효과적으로 실행되고 지속성을 유지하기 위해서는 DMP 작성 지원, 데이터 관리와 표준화 지원, 데이터 관리·공유 인프라 등 연구자에게 다양한 지원 여건이 필요하다. 또한 DMP 가이드라인과 이행 절차는 국가 또는 기관 차원의 연구데이터 관리·활용을 위해 필수적이다. DMP 정책을 적극적으로 시행하고 있는 해외와 달리 국내에서는 연구자와 연구기관에서 활용 가능한 DMP 사례나 지침, 수행 경험 등이 아직 부족한 상황이다. 이 연구에서는 국내의 DMP 추진 현황과 이슈를 살펴보고, 국내에서 DMP 이행을 위해 공통적으로 요구되는 구성 요소와 체크리스트, 적용프로세스, 그리고 연구기관의 DMP 이행 방안을 종합적으로 제시한다.

II. 관련 연구

김지현의 연구(2013)에서는 미국, 영국, 캐나다, 호주 의 15개 정부 연구 펀딩 기관에서 제공하는 데이터 관 리 정책을 분석하였다[7]. 15개 분석대상 해외 기관들 중 11개 기관에서 DMP 정책을 제시하고 있으며, 각 기관의 DMP 구성요소를 비교하였다.

한중규의 연구(2015)에서는 한국지질자원연구원의 DMP 양식을 제시하고 시범 수집 결과를 분석하고 있 다[9]. 연구자가 작성하는 데이터에 대한 상세 정도의 차이가 심함을 지적하고 데이터의 종류, 내용, 양 등에 대한 기준 마련의 필요성을 언급하고 있다.

구찬미의 연구(2019)에서는 연구기록물 관리의 관점 에서 DMP의 효용성을 분석하고, 연구기록물 관리 계 획(RRMP)을 적용할 것을 제안하였다[10]. RRMP에서 는 연구기록물의 특성을 반영하기 위하여 '계획서정보', '연구개발사업정보', '연구기록물정보'의 세 영역으로 구 성되어 있다.

Simms의 연구(2016)에서는 DMP 정책과 더불어 대 표적인 DMP 작성 도구인 DMPOnline, DMPTool의 현황에 대해 소개하였다[11]. 또한 이러한 DMP의 작 성 지원은 연구기관에서의 연구데이터 관리 서비스의 핵심 요소로 인식되고 있다[12-15]. 특히 해외의 경우 많은 대학과 연구기관에서 데이터 전문가와 IT 전문인 력, 주제별 사서를 중심으로 연구자의 DMP 작성과 연 구데이터 관리·공유를 위해 지원하고 있다.

연구자가 작성하여 제출된 DMP는 연구데이터에 대 한 다양한 정보를 포함하고 있어 체계적으로 관리·활용 되어야 하며, 연구개발과제, 연구데이터, 연구자, 논문 등 다른 연구개발정보 또는 성과물과의 연계가 필요하 다. Simms의 연구(2017)에서는 DMP의 관리·활용연 계를 위하여 기계실행 가능한 DMP 서비스 (Machine-Actionable DMP, maDMP)를 제안하고 있다[16]. 또한 RDA DMP Common Standards WG 에서는 maDMP 구현을 위한 DMP 데이터 모델(DMP Common Standards)을 제시하고 있다[17].

선행 연구에서는 주로 연구데이터 관리 정책과 서비 스의 부분으로써 DMP의 구성요소를 중심으로 분석하 였다. 이 연구에서는 국내외의 연구자와 연구기관에서

DMP 이행을 위한 필요 요소와 적용 프로세스, 그리고 연구기관의 추진 방안을 제시하고자 한다.

III. 데이터 관리 계획

1. 개요

데이터 관리 계획(Data Management Plan, 이하 DMP)은 연구데이터의 생산·보존·관리 및 공동활용 등에 관한 계획을 말한다(공동관리규정 제2조 19항). 국가연구개발과제 제안시 계획서와 함께 2~3쪽 분량의 별도 문서로 제출하는 경우가 일반적이다.

DMP 정책을 폭넓게 시행하고 있는 해외에서는 연구 자들에게 데이터의 관리와 공유 인식을 고취시키는 중 요한 정책으로 인식되고 있다. 연구데이터는 연구 진행 기간 중에 체계적으로 관리되지 않으면 품질을 보장할 수 없고 손실될 가능성이 높다. DMP는 연구 수행 전에 데이터에 대한 생산, 관리, 공유, 활용에 대한 계획을 수립하게 함으로써 데이터의 품질과 더불어 재사용을 높 일 수 있는 효과적인 방안이 될 수 있다. 또한 국가 또 는 기관 차원에서 수행된 연구개발과제로부터 생산되 는 데이터 현황을 파악하는 데 핵심적인 역할을 하게 된다.

국가연구개발사업의 범부처 총괄 규정인 공동관리규정이 연구데이터의 체계적인 관리를 위하여 [표 1]과 같이 개정되었으며, 2019년 9월에 시행되었다. 다만 DMP의 제출은 중앙행정기관의 장이 필요하다고 인정 하는 연구개발과제에만 해당하기 때문에 한계가 있으 며, 기존의 9대 성과물처럼 관리하는 전담기관이나 관 리시스템은 아직 존재하지 않은 상황이다.

표 1. 공동관리규정 주요 개정 내용

정의 및 도입 근거	<ul style="list-style-type: none"> 연구데이터 정의(2조 18항 신설) DMP 정의(2조 19항 신설) 연구데이터 생산·보존·관리 및 공동활용 시책 수립·추진 (25조 28항 신설)
제출 의무사항	<ul style="list-style-type: none"> 연구개발계획서 작성 시 DMP 제출 요구(6조 4항 개정) 연구개발과제의 선정 시 검토 요구(7조 3항 개정) 과제 협약 체결 시 DMP 제출 요구 및 준수(9조 1항 개정)
보고	<ul style="list-style-type: none"> 최종보고서 작성 시 DMP 제출 추가 요구(15조 2항 개정)
관리 및 연계	<ul style="list-style-type: none"> DMP에 포함된 정보 등록 및 연계(25조 16항 개정)

연구기관 또는 연구자들이 국가연구개발사업에서 참조하여 활용할 수 있도록 ‘국가연구개발사업 연구관리 표준매뉴얼’에 연구데이터와 DMP 관련 내용을 반영했다. DMP에 대한 표준 서식을 제시하고 있으며, 한국연구재단과 정보통신기획평가원의 DMP 예시가 수록되어 있다. 다년 사업인 경우 연구데이터에 대한 연차 실적 보고 양식도 제시하고 있다.

국가과학기술연구회의 ‘소관연구기관 주요사업 운영 규정’ 또한 공동관리규정에 따라 [표 2]와 같이 개정되어 2019년 12월 3일 시행되었다. 공동관리규정과 마찬가지로 DMP의 제출은 연구기관의 장이 필요하다고 인정하는 연구개발과제에 한정하고 있으며, DMP 제출은 공동관리규정을 준용하여 연구기관에서 자율적으로 규정하도록 하고 있다.

표 2. 소관연구기관 주요사업 운영규정 개정 내용

정의 및 도입 근거	<ul style="list-style-type: none"> 연구데이터 정의(2조 5항 신설) DMP 정의(2조 6항 신설) 연구데이터 관리·활용 의무(4조 4항 개정)
DMP 제출	※ 공동관리규정을 준용하여 연구기관에서 자율적으로 규정
DMP 보고	<ul style="list-style-type: none"> 최종보고서 작성 시 DMP 이행여부 제출 추가 요구(11조 2항 개정)

2. 구성 요소

DMP는 [그림 1]과 같이 연구자가 연구기획 시 수집 또는 생산 예정인 연구데이터에 대한 개요, 관리 방안, 공유 방안, 관리 책임자 등 개략적인 내용을 담는 항목으로 구성되어 있으며[18], 구성요소는 국가별 또는 기관별로 매우 다양하다[7].

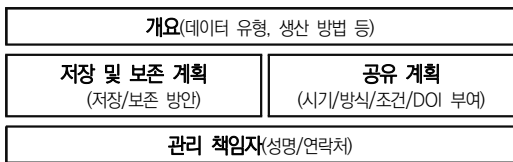


그림 1. DMP 구성 요소

이 연구에서 제안하는 DMP 양식(안)은 [표 3]과 같이 구성될 수 있으며, 연구기관별로 필요한 요소를 추가하여 활용한다. 연구데이터가 생성되지 않는 경우에는 “데이터가 생산되지 않음”을 표기하도록 한다.

표 3. DMP 양식(안)

연구과제 개요	연구과제명, 수행기간, 주관기관, 책임자, 연구개요 등 기재
(1) 연구데이터 개요	
데이터 설명	<ul style="list-style-type: none"> 데이터명과 간략한 설명* <ul style="list-style-type: none"> * 연구 노트, 예비 분석, 과학 논문 초안, 장래 연구 계획, 동료 평가 또는 동료와의 커뮤니케이션, 실험물(physical objects) 등은 제외 데이터의 유형*, 포맷** <ul style="list-style-type: none"> * 텍스트, 이미지, 음향, 동영상, 수치 등 ** CSV, JPG, HDF, XML 등 데이터의 크기, 건수 데이터 및 메타데이터 관련 표준* <ul style="list-style-type: none"> * 해당 분야에서 활용되는 표준이 없는 경우 데이터 기술에 필요한 주요 항목 기재
생산 수집 방법	<ul style="list-style-type: none"> 연구데이터 생산 및 수집 방법 ※ 사용된 장비 및 소프트웨어 등 기재
(2) 연구데이터의 저장 및 보존 계획	
데이터 저장	<ul style="list-style-type: none"> 연구 수행 중 데이터 저장 및 관리 방안 ※ 연구데이터 저장소 기재
연구 종료 후 데이터 보존	<ul style="list-style-type: none"> 연구 종료 후 장기적인 데이터 보존 계획 ※ 연구데이터 보존 장소와 필요 기간 기재 ※ 연구자의 개인 컴퓨터, 연구실 서버 등은 장기 보존 방법으로는 적합하지 않으며 전문 데이터센터나 데이터 리포지토리를 활용하거나 소속 연구기관이 제공하는 안정적인 데이터 저장소를 권장함
(3) 연구데이터의 공유 계획	
데이터 공유	<ul style="list-style-type: none"> 연구데이터 공개 시기 <ul style="list-style-type: none"> ※ 논문 발표 등을 위해 과제 완료 후 최대 2년 이내의 공개 유예 기간 설정 가능 연구데이터 공개 방법 <ul style="list-style-type: none"> ※ 공개에 필요한 자원(소프트웨어, 장비) 기재 데이터 식별자(DOI 등) 정보 연구데이터 접근, 이용 조건 <ul style="list-style-type: none"> ※ Creative Commons License(CCL), Open Data Commons(ODC), Korea Open Government License(KOGL) 등 표준 라이선스 중 적절한 이용허락 조건을 선택하여 적용하거나, 별도의 조건을 기재
공개 공유 제한	<ul style="list-style-type: none"> 연구데이터 공개·공유 제한 사항 ※ 보안과제, 개인정보보호 등으로 인해 공개가 어려운 경우 비공개 사유 기재
(4) 연구데이터의 관리 책임자	
책임자	<ul style="list-style-type: none"> 데이터의 관리, 저장, 공유에 대한 세부 사항을 점검하고 이행할 연구자의 이름, 소속기관, 연락처 등 기재 ※ 복수 책임자 지정 가능

DMP 작성 예시는 [표 4]와 같다.

표 4. DMP 예시

연구과제 개요	<ul style="list-style-type: none"> 연구과제명 : 000 수행기간 : 2019.1.1. ~ 2019.12.31 주관기관 : 000연구원 책임자 : 000 연구개요 : 본 연구과제는 스마트 시티 구현을 위한 관련 연구데이터를 수집, 관리하는 시스템을 개발함. 또한, 다양한 스마트 시티 데이터를 통합 관리하고 검색하는 시스템을 개발하며, 해당 데이터를 분석하고 모니터링하는 기술도 개발하여 스마트 시티 관련 데이터를 활용하는 생태계를 구현하는 것이 목표임
(1) 연구데이터 개요	
데이터 설명	<ul style="list-style-type: none"> 데이터명, 유형, 포맷, 용량 <ul style="list-style-type: none"> - 대기질 측정 데이터, 수치, CSV/JSON, 100MB - 도로영상 데이터, 동영상, MPEG4, 10GB 데이터 관련 표준 <ul style="list-style-type: none"> - 대기질 측정 데이터 : CSV / JSON - 도로영상 데이터 : MPEG-4 메타데이터 관련 표준 <ul style="list-style-type: none"> - 자체 메타데이터 활용 - 대기질 측정 메타데이터 : 차량ID, 일시, 위도, 경도, PM25, PM10, NO2, SO2, VOC 등 - 도로영상 메타데이터 : 차량ID, 일시, 저장화일(위치), 파일사이즈 등
생산·수집 방법	<ul style="list-style-type: none"> 연구데이터 생산 및 수집 방법 <ul style="list-style-type: none"> - 대기질 측정 데이터 : 대기질 측정 센서 장비로부터 이동 차량 2대에 탑재하여 하루 8시간 측정하여 생성하고 무선통신을 사용하여 서버에 저장 - 도로영상 데이터 : GoPro 센서를 활용하고 이동 차량 2대에 탑재하여 하루 8시간 측정하여 생성하고 내부 저장장치에 저장 사용된 소프트웨어 및 장비 <ul style="list-style-type: none"> - 자체 제작 장비
(2) 연구데이터의 저장 및 보존 계획	
데이터 저장	<ul style="list-style-type: none"> 연구 수행 중 데이터 저장 및 관리 방안 <ul style="list-style-type: none"> - 연구 부서 내부 서버를 활용하여 저장 관리하고 별도의 백업 스토리지로 정기적 백업 실행
연구 종료 후 데이터 보존	<ul style="list-style-type: none"> 연구 종료 후 장기적인 데이터 보존 계획 <ul style="list-style-type: none"> - 기관 리포지터리에 등록하여 장기 보관하여 관리 연구데이터 보존 필요 기간 <ul style="list-style-type: none"> - 최소 10년 이상 보존 필요
(3) 연구데이터의 공유 계획	
데이터 공유	<ul style="list-style-type: none"> 연구데이터 공개 시기 <ul style="list-style-type: none"> - 연구 종료 후 즉시 공개 연구데이터 공개 방법 <ul style="list-style-type: none"> - 기관 리포지터리를 통해 데이터 개방 및 공유 데이터 식별자 정보 <ul style="list-style-type: none"> - DOI 부착 예정 연구데이터 접근 및 이용 조건 <ul style="list-style-type: none"> - 저작자표시-비영리-동일조건 변경 허락(BY-NC-SA) - 연구목적인 경우 무료로 활용하고, 상업용으로 활용할 경우 별도의 계약 또는 데이터 이용 협약을 통해 사용
공개·공유 제한	<ul style="list-style-type: none"> 연구데이터 공개·공유 제한 사항 <ul style="list-style-type: none"> - 해당 없음
(4) 연구데이터의 관리 책임자	
책임자	<ul style="list-style-type: none"> 000, 000연구원, 000부서, 전화번호, 이메일

3. 적용 프로세스

연구계획서는 대부분 관련 연구자들과의 협력이 필요하므로, DMP는 해당 연구 분야의 실무 커뮤니티가 개발한 절차, 표준과 모범 사례를 고려해야 한다. DMP 검토는 연구계획의 합리성, 적합성, 영향력 모두를 평가하기 위해 중요한 사항이다. 과제 완료 후 연구데이터에 대한 접근과 공유는 IRB, 개인정보보호, 기밀유지, 데이터 보안, 지적 재산에 대한 적절한 보호가 이루어져야 한다. 또한 정책이나 규정에 의해 달리 제한되지 않는 한, 가능한 빠른 시일내에 데이터를 접근과 공유가 이루어질 필요가 있다.

DMP 정책이 효과적으로 실행되고 지속성을 유지하기 위해서는 연구자에게 다양한 지원이 필요하다. 연구자들을 대상으로 DMP 작성에 대한 직접적인 지원과 질문에 대한 답변, 컨설팅을 할 수 있는 체계가 우선적으로 필요하다. 또한 향후 연구데이터가 효과적으로 수집, 보존, 공유되려면 데이터의 포맷, 메타데이터 그리고 데이터 관리 프로세스 등이 표준화되어야 한다. 연구자들이 직접 수집, 생산한 연구데이터를 기탁, 저장할 저장소를 포함한 인프라의 구축이 뒤따라야 한다. 따라서 초기 단계에는 DMP의 제출과 확인을 시행하고, 데이터 공유에 필요한 규정, 인프라, 전문 인력 지원 등 협력체계가 체계적으로 준비되어야 하며, 각 이해관계자들에게 충분한 준비 기간을 부여해야 한다.

연구과제 기획·선정 단계에서는 DMP 작성, 검토 과정이 추가된다. 연구자에게 DMP 작성 지침과 도구가 제공되어야 하며, 연구자들은 과제 기획시 생산, 수집될 연구데이터에 대한 관리·공유 계획을 수립하여 제출하게 된다. 제출된 DMP는 과제 선정 평가 시 검토되어야 하며, 필요시 피드백과 추가 심의가 이루어질 수 있다. 또한, 한국연구재단, 정보통신기획평가원 등 연구관리 전문기관은 과제지원시스템에 DMP에 포함된 연구개발정보를 포함하여야 한다.

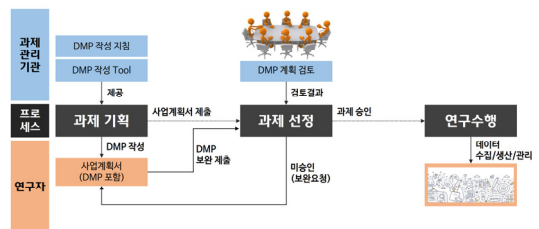


그림 2. 연구과제 기획·선정 단계의 DMP 작성 및 검토

연구과제 수행 중에 DMP가 변경될 경우는 연구과제 종료 시점인 최종보고서 작성 시에 DMP에 대한 갱신과 이행에 관한 사항을 포함하여 보고해야 하며, 다년도 계속사업의 경우 초기 DMP에 반영하지 못한 연구데이터 관련 실적을 연차실적·계획서에 포함해서 작성, 보고하도록 한다.

연구과제 수행·평가 단계에서는 DMP 이행여부 점검에 대한 추가 심의가 수행되며, 점검 결과 미흡인 경우 폐널티 또는 추가 조치가 이루어져야 한다. 또한 공개제한 사항이 없을 경우 데이터 공유가 이루어질 수 있도록 유도해야 한다.

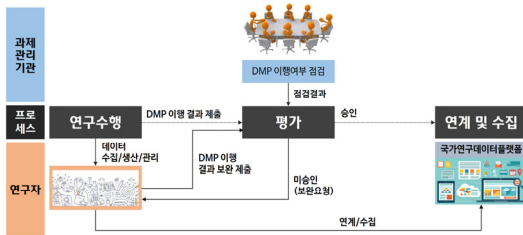


그림 3. 연구과제 수행·평가 단계의 DMP 이행여부 점검

4. 체크리스트

연구과제의 선정 시 연구계획서의 부속문서로 제출된 DMP를 검토하게 된다. 연구계획서 평가 시 해당 분야의 전문가 이외에 데이터 전문가의 참여가 바람직하다. 해당 분야의 전문가들이 그 분야의 데이터 생산과 공유에 대해서는 잘 알고 있었지만, 연구데이터의 표준화, 문서화, 상호운용성 등을 높이기 위해 대규모의 데이터 생산이 예상되거나 데이터의 공유가 필수적인 분야의 연구계획서 평가에는 별도의 데이터 전문가의 참여가 요구된다.

DMP의 심사가 평가 접수에 포함되지 않는 경우 내용이 매우 부실하거나 해당 분야에서 인정하는 범위와 차이가 있는 경우에는 DMP의 보완을 요구할 수 있다. 이에 대한 구체적인 절차가 향후 평가에 포함될 필요가 있다.

DMP 심사 과정에서 장기적인 보존과 공유의 가치가 큰 데이터의 수집이 예상되는 연구계획서는 ‘추적’을 요청하도록 하여 향후 해당 데이터에 대해 효과적인 관리와 공유를 촉진할 수 있다.

DMP의 전반적인 심사 원칙과 체크리스트는 아래와 같다.

- 계획은 명료하고 구체적이어야 한다.
- 해당 연구 분야 혹은 연구문제에 대한 표준적이거나 모범적인 관행을 따라야 한다.
- 계획은 현실적이어야 하며, 데이터 공유를 적극적으로 지향해야 한다.

표 5. DMP 체크리스트(안)

데이터 개요	<p>생산 또는 수집하고자 하는 데이터의 설명</p> <ul style="list-style-type: none"> • 데이터 생산 및 수집 방법이 해당 분야의 통상적인 규범이나 표준적인 방법을 따르고 있는가? • 연구 논문의 표나 그림, 보고서의 부록에 수록되는 내용, 예비 분석결과, 논문이나 저술의 초안, 연구노트, 보고서 등은 해당되지 않는데 이런 유형의 데이터가 제시되고 있는가? • 데이터 포맷 등이 가능한 범위에서 공개된 표준을 따르고 있는가?
데이터의 저장 및 보존 계획	<p>데이터의 저장 및 장기적인 보존 계획</p> <ul style="list-style-type: none"> • 구체적인 데이터의 저장 및 장기 보존 계획이 제시되고 있는가? • 데이터 저장 방법이나 저장소는 안전성이 있는가? • 데이터 저장 및 장기 보존 계획이 실현 가능한가?
데이터 공유 계획	<p>데이터의 공개 및 공유 계획 작성</p> <ul style="list-style-type: none"> • 연구데이터의 공유 계획이 작성되었는가? • 공유 계획이 구체적이고 실현 가능한가? • 해당 분야의 통상적인 규범이나 표준적인 방법으로 공유 계획이 작성되었는가? • 데이터 공유에 제한 사항(개인정보, 기밀 유지 필요, 지적재산권 및 잠재적인 상업적 가치 보호 등) 등이 파악되고 그에 대한 대안이 제시되고 있는가? • 데이터 사용에 대한 제한 사항(라이선스 조건, 엠바고, 공유 제한 사항)이 파악되고 있는가?
데이터 관리 책임자	<p>데이터의 관리 및 공유 책임자 기재</p> <ul style="list-style-type: none"> • 데이터의 전반적인 관리, 저장, 공유와 관련된 세부 사항을 이행할 연구자의 이름, 연락처가 기재되어 있는가? (복수의 관리 책임자 기재 가능)

DMP의 기재 내용에 대해서는 연차보고, 최종보고 등을 통해 이행 여부를 확인해야 하며, 연구책임자에게 점검 결과가 전달되고 향후 연구계획서 평가에 반영될 수 있어야 한다. 데이터의 중요성 혹은 가치로 인해 심사과정에서 ‘추적’으로 식별된 연구계획서는 관리 부서 지정과 별도의 이행여부 검토 절차가 수행되어야 한다. 앞서 제출된 DMP에 대한 이행이 이루어지지 않거나 미흡한 경우, 후속 연구과제에 대한 참여 시 선정 평가에 반영되어야 한다.

5. 국내 추진 현황

2019년 과기정통부 산하 기관인 한국연구재단과 정보통신기획평가원은 일부 연구개발과제에 DMP 제출을 시범 적용했다. 한국연구재단은 2개 분야 8개 사업, 정보통신기획평가원은 10개 과제에 DMP를 제출하도록 했으며, 매년 적용 대상이 확대될 예정이다. 국가과학기술연구회 산하 정부 출연 연구기관은 ‘소관기관 주요사업 운영규정’ 개정에 따라 개별 기관의 주요사업 관리규정을 개정 중에 있다. 한국지질자원연구원, 한국한의학연구원, 한국과학기술정보연구원 등은 DMP 제출 규정을 시행하고, 연구데이터 관리 인프라를 운영 중이며, 기관별로 차이는 있으나, 많은 기관들이 2021년 주요사업의 일부 과제에 대해 DMP 제출을 시범 적용할 예정이다.

한국연구재단에 제출된 시범과제의 DMP를 분석한 결과 연구자들은 아직 DMP 작성에 낯설어 하고 있는 상황이다. 먼저, 연구노트, 문서, 소프트웨어 등을 포괄하여 연구데이터로 기술하고 있다. 연구데이터는 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터”로 정의되며(공동관리규정 2조 18항), 연구 노트, 예비 분석, 과학 논문 초안, 장래 연구 계획, 동료 평가 또는 동료와의 커뮤니케이션, 실물(Physical Objects) 등은 제외된다[19]. 둘째, 연구데이터에 대해 작성된 내용이 모호하거나 연구내용과 혼재되어 기술되고 있다. 데이터명, 크기 또는 건수, 유형, 포맷, 데이터 표준 등이 보다 구체적으로 작성될 필요가 있다. 셋째, 데이터는 대부분 논문이나 특허 확보 후에 공개할 계획이며, 방식도 논문 등을 통해 공개하는 경우도 많다. 데이터의 공개는 그 활용이 유의미할 수 있도록 최대한 신속하게 이루어져야 한다. 보안과제, 개인정보보호 등으로 인해 공개가 어려운 데이터를 제외하고는 공개시기를 과제 완료 후 최대 2년 이내로 제한하고, 시스템을 통해 공개하도록 권장되어야 한다.

전반적으로 아직 연구데이터 공유·활용을 위한 한계가 존재한다. 국가 연구개발사업은 각 부처별로 서로 다른 법령과 관리지침에 근거하여 추진되고 있으며, 국가 연구개발과제의 공통규범인 공동관리규정도 대통령령

으로 제정, 운영되고 있어 그 규범적 효력이 미흡하다. 공동관리규정은 국가 예산이 투입되는 부처별 연구개발사업에 해당하나, DMP 제도는 전면 시행이 아닌 중앙행정기관의 장이 인정하는 과제에 한한다. 또한, 부처별 연구개발사업 관리 규정에 대한 예외 사항도 존재한다.

국내의 연구자들 역시 생산된 연구데이터를 연구자 본인의 소유로 인식하는 경향이 있으며, 공유 경험이 부족한 상황이다. 연구데이터와 관련한 인프라와 지원의 부재로 인해 국가 또는 기관 차원에서 연구데이터를 공유하고 활용하기에는 한계가 있다. 많은 연구자들이 다른 연구자들이 생산한 연구데이터를 활용하기 어려운 가장 큰 이유를 데이터의 존재와 위치를 모르고 접근이 어려워서라고 답하고 있다.

IV. 연구기관의 DMP 이행 방안

연구데이터 공유·활용 활성화를 위해서는 연구자들이 소속되어 있는 연구기관에서의 DMP 이행이 우선적이며, 밑바탕이 되어야 한다. 따라서 기관 차원에서 연구데이터 공유와 활용을 위한 체계적인 정책 마련과 그 정책을 뒷받침하는 가이드라인을 제공하여 연구자들이 이해하고 참여할 수 있는 체제 마련이 중요하다.

기관 차원에서 연구데이터 관리·활용 체제를 구축하기 위해서 다음과 같은 단계의 프로세스를 추진한다.



그림 4. 연구데이터 관리·활용 체제 구축 프로세스

현황 조사 단계에서는 연구자들이 생산한 연구데이터 현황을 조사하고 연구데이터를 이해하고 향후 관리·공유 체계 구축의 기초 자료를 확보하기 위해 다음과 같은 영역을 조사·분석한다.

표 6. 현황 조사 단계

구분	주요 내용
조사 방법	· 설문조사와 FGI(Focus Group Interview)를 병행
조사 영역	· 데이터 생산 과제 · 데이터 생산 연구자 · 데이터 유형 및 용량, 생산방식 · 데이터 활용 형태 및 분야
산출물	· 데이터 생산·관리·활용 현황 · 관리 대상 데이터 선정 · 메타데이터 표준 및 주요 항목 · 데이터 분류 체계 · 데이터 활용 기술 및 도구 · 인프라 선정 기준

요구조사·분석 단계에서는 각 이해 관계자(과제 관리자, 연구자, 데이터 관리자 등)들의 요구사항을 조사하여 연구데이터 생태계 구축을 위해 다음과 같은 영역을 조사·분석한다.

표 7. 요구조사 및 분석 단계

구분	주요 내용
조사 방법	· 1단계 조사 결과를 토대로 인터뷰를 실시하여 개별 이해관계자(과제 관리자, 연구자, 데이터 관리자 등)의 세부적인 요구 사항을 분석
조사 영역	· 성과물 관리 요구사항 · 데이터 등록 요구사항 · 데이터 관리 요구사항 · 데이터 활용 요구사항
산출물	· 데이터 관리·활용 요구사항 · 개별 이해관계자들의 역할과 책임 · 연구데이터 관리 및 활용 프로세스 · 연구데이터 리포지터리 기능 및 서비스

거버넌스 체계 구축 단계에서는 데이터 공유 정책, DMP, 평가 등 가이드라인을 포함하여 기관 차원의 연구데이터 관리와 공유를 위한 거버넌스 체계를 구축하게 된다. 기관 차원에서 연구데이터를 관리하고 활용하기 위해서는 다양한 이해 관계자가 참여하는 전사적인 태스크 포스를 구성하여 추진하는 것이 효과적이며, 다음과 같은 산출물을 고려할 필요가 있다.

표 8. 거버넌스 체계 구축 단계

구분	주요 내용
수행 방식	· 다양한 이해 관계자가 참여하는 전사적 태스크 포스를 구성하여 추진
산출물	· 연구데이터 공유 정책 · 기관 연구개발사업 관리 규정(연구데이터, DMP 제출 등) · 연구데이터 관리 및 공유 가이드라인 · 연구데이터 생태계 활성화 방안(평가 및 인센티브 등)

R&R 및 이행방안 마련 단계에서는 연구데이터 생태계를 이루는 연구 관리자, 연구자, 데이터 관리자 등 이해관계자들의 역할과 책임을 명시하고, 해당 역할을 수행하는 데 필요한 내부 규정과 지침을 마련한다.

표 9. R&R 및 이행방안 마련 단계

구분	주요 내용
역할과 책임 정의	· 각 이해관계자들의 역할과 책임 정의 (예시) - (연구자) DMP 제출과 데이터 공유 - (과제 관리자) 연구과제 계획서와 DMP를 검토하고 연구 수행 후 최종 단계에서 DMP 이행여부를 점검 - (데이터 관리자) 제출된 연구데이터의 체계적인 관리를 수행
세부 지침 마련	· 연구 프로세스 정의 및 관련 규정 개정 · 데이터관리계획(DMP) 서식 및 작성 가이드라인 · 데이터관리계획(DMP) 체크리스트 · 연구데이터 등록 프로세스 및 지침서

마지막 연구데이터 관리·공유 시스템 구축 단계에서는 기관 차원에서 연구데이터를 관리하고 공유하는 리포지터리 시스템을 구축하고 운영한다. 데이터 관리 부서에서 개별 이해 관계자들의 요구사항을 수렴하여 필요한 서비스와 기능을 정리하고, 서비스와 기능의 요구사항을 충족하는 오픈 소스 시스템을 도입하여 연구데이터 리포지터리 시스템으로 커스터마이징하게 된다. 시스템 구축 시 이용자에 대한 서비스 정책과 시스템을 운영하기 위한 세부적인 가이드라인을 마련해야 하며, 체계적인 서비스를 운영을 위해 제공하는 서비스에 대한 모니터링과 피드백 체계도 필요하다.

표 10. 시스템 구축 및 운영 단계

구분	주요 내용
시스템 도입 검토	· 서비스 요구사항 분석 · 서비스 기능 정의 · 오픈 소스 기반 시스템 선정 · 커스터마이징 방안 마련
시스템 구축 및 운영	· 서비스 정책 마련 · 서비스 운영 가이드라인 마련 · 서비스 모니터링 및 피드백 체계 마련

V. 결론

연구자의 연구 환경이 데이터 기반의 디지털 환경으로 전환되고 있으며, ICT 기술과 연구 장비의 발달로

많은 연구데이터가 생산되고 있다. 연구데이터의 공유와 활용은 연구의 효율성을 높이는 동시에 투명한 연구 프로세스를 조성하여 연구의 재현성과 신뢰성을 확보하는 데 중요한 역할을 하게 된다. 주요 선진국에서는 공적 지원에 의해 수행된 연구과제에서 생산되는 데이터의 공유·활용을 위한 정책과 인프라 구축이 적극적으로 추진되고 있다. 국내에서도 오픈 연구데이터 추진을 위해 2019년도에 관련 규정이 개정되어 DMP 제도가 시행되고 있다.

이 연구에서는 DMP의 국내 추진 현황과 이슈를 살펴보고, DMP의 공통 양식과 체크리스트, 연구기관에서 DMP 이행을 위한 방안을 제안했다. 연구데이터는 분야별로 그 특성과 유형이 매우 다양하며, 그 활용 방법 또한 상이하다. 따라서 기관별, 분야별 특성을 고려하여 조정 또는 구체화할 필요가 있다. 또한, DMP의 체계적인 관리와 활용을 위하여 향후 DMP의 데이터 모델과 기계 실행가능한 DMP 서비스 구현·연계 방안에 대한 연구를 진행할 계획이다.

DMP 제도는 연구데이터 공유와 활용의 시작이다. 연구자가 제출하는 DMP는 연구 활동에서 생산 또는 수집되는 데이터를 자발적으로 개방하고 공유하는 새로운 연구 문화로의 전환을 위한 기반이 된다. 많은 연구자들이 누가 생산했는지, 어느 곳에 저장되었는지, 공개·활용 가능한지에 대한 정보가 없어서 기존의 연구데이터를 활용하기 힘들다고 어려움을 토로하고 있다. DMP 제도가 연구 현장에 정착될 수 있도록 연구자의 권리를 보호하고 평가에도 반영할 필요가 있다. 기존 연구개발 결과에 대한 평가는 논문, 특허, 연구보고서 실적을 중심으로 이루어지고 있어, 연구자가 데이터 저널 등을 통해 연구데이터를 출판하면 기존 논문 저널에 출판한 실적과 동일하게 인정하고, 연구데이터의 공유·활용 실적을 평가에 반영해 주는 제도를 고려해야 한다. 연구자의 연구데이터 관리·공유 인식 전환을 위한 전략과 지원을 지속적으로 추진해야 하며, 연구데이터의 접근성 향상과 자발적인 공개·공유 활성화를 위해 연구데이터 리포지터리, 분야별 데이터센터 등 기반 인프라 구축과 운영이 시급하다.

참고 문헌

- [1] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 2009.
- [2] OECD, "Making Open Science a Reality," OECD Science, Technology and Industry Policy Papers, No.25, 2015.
- [3] The Royal Society, "Science as an open enterprise," The Royal Society Science Policy Centre Report, 2012.
- [4] 신은정, "오픈 사이언스(Open Science)에 관한 OECD 논의 동향과 시사점," STEPI 동향과 이슈, 제 22호, 2015.
- [5] J. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research," White House OSTP Memorandum, 2013.
- [6] Open Research Data Forum, "Concordat on Open Research Data," Open Research Data Forum Report, 2016.
- [7] 김지현, "국의 정부연구비지원기관의 연구데이터 관리 정책 분석 - 미국, 영국, 캐나다, 호주를 중심으로," 한국문헌정보학회지, 제47권, 제3호, pp.251-274, 2013.
- [8] 최명석, "오픈 연구데이터 정책 동향," IE 매거진, 제 24권, 제4호, pp.30-35, 2017.
- [9] 한종규, *KIGAM 지질자원정보 통합관리체계 구축*, 한국지질자원연구원 보고서, 2015.
- [10] 구찬미, 김순희, "과학기술분야 연구기관의 DMP를 적용한 연구기록물 관리," 한국기록관리학회지, 제19권, 제1호, pp.1-21, 2019.
- [11] S. Simms, S. Marisa, J. Sarah, and R. Marta, "The Future of Data Management Planning: Tools, Policies, and Players," *International Journal of Digital Curation*, Vol.11, No.1, pp.208-217, 2016.
- [12] 김지현, "대학도서관의 연구데이터관리서비스에 관한 연구 - 미국 연구중심대학도서관을 중심으로," 한국비블리아학회지, 제25권, 제3호, pp.165-189, 2014.
- [13] 심원식, "미국 대학도서관의 연구데이터 지원 서비스 사례 연구," 한국문헌정보학회지, 제50권, 제4호, pp.311-332, 2016.
- [14] 김주섭, 김선태, 최상기, "연구 데이터 관리 및 서비스

스를 위한 핵심요소의 기능적 요건,” 한국문헌정보학회지, 제53권, 제3호, pp.317-344, 2019.

[15] A. Cox, M. Kennan, E. Lyon, S. Pinfield, and L. Sbaffi, “Progress in Research Data Services : An International Survey of University Libraries,” International Journal of Digital Curation, Vol.14, No.1, pp.126-135, 2019.

[16] S. Simms, S. Jones, D. Mietchen, and T. Miksa, “Machine-actionable data management plans (maDMPs),” Research Ideas and Outcomes 3, e13086, 2017.

[17] <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

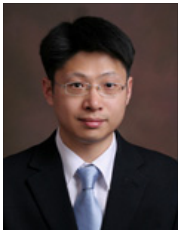
[18] S. Jones, “How to Develop a Data Management and Sharing Plan,” DCC How-to Guides. Edinburgh: Digital Curation Centre. 2011.

[19] 윤종민, “과학데이터의 공유 및 활용 촉진을 위한 법적 과제,” 법학연구, 제27권, 제1호, pp.597-625, 2016.

저 자 소 개

최 명 석(Myung-seok Choi)

정회원



- 2005년 8월 : KAIST 전산학과(공학박사)
- 2005년 6월 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 선임연구원

〈관심분야〉 : 오픈사이언스, 연구데이터관리, 인공지능 등

이 상 환(Sanghwan Lee)

정회원



- 2004년 8월 : 고려대학교 S/W공학과(공학석사)
- 1995년 4월 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 책임연구원

〈관심분야〉 : 연구데이터, 빅데이터, 오픈사이언스, 딥러닝 등