

# DMP(Data Management Plan), 연구데이터 공유와 활용의 시작

이상환  
(KISTI 연구데이터공유센터)

## 1

### 개요

#### 가. 배경 및 필요성

국가 차원에서 연구데이터 공유와 활용을 위한 배경은 크게 두 가지로 볼 수 있다. 첫째, 4차 산업혁명의 도래로 데이터의 중요성이 그 어느 때보다도 증대되었고, 데이터의 활용이 과학기술 경쟁력을 넘어 국가 경쟁력을 좌우하는 핵심으로 대두되었다. 다양한 기술의 개방과 융합을 핵심으로 하는 4차 산업혁명은 산업구조 및 시장경제 모델에 가져올 파괴적인 혁신이며, 데이터 혁명에 기반을 두기 때문이다). 점차적으로 4차 산업혁명 시대에는 데이터가 중요해지면서 데이터 활용의 역량이 국가 과학기술 경쟁력의 핵심으로 대두되고 있다.

연구자의 연구 환경이 디지털 기술의 발전으로 인해 데이터 중심으로 전환되고, 연구 성과를 확산시키는 방식과 연구를 수행하는 방식 자체를 변화시키고 있다<sup>2)</sup>. 현대 과학은 연구개발 사이클에서 데이터의 생성, 처리, 분석 등 데이터 중심으로 전환되어 연구데이터의 품질과 양에 따라 연구 결과의 품질도 좌우할 수 있다.

둘째, 전 세계가 연구데이터를 기반으로 개방형 혁신체제를 추구하고 있다. 오픈 액세스 저널과 셀프아카이빙 플랫폼을 통한 온라인 방식으로 논문을 공개와 공유하는 사례가 증가하고 있다<sup>3)</sup>. 더욱이, OECD, 미국, 영국 등 선진국은 연구개발 과정과 연구 성과를 국가 차원에서 개방과 공유하는 오픈사이언스 정책이 확산되고 있다. 이러한 연구데이터의 개방과 공유는 국가 연구개발의 투자 효율성, 연구 결과의 투명성 확보, 빠른 지식공유 등 사회경제적 측면에서도 긍정적 효과를 입증<sup>4)</sup>하고 있다. 공개된 연구 결과물(논문, 데이터, 소프트웨어 등)은 재사용하여 불필요한 반복적인 실험을 하지

1) 김진하 (2016), 「제4차 산업혁명 시대, 미래사회 변화에 대한 전략적 대응 방안 모색」, KISTEP In1 제15호, 한국과학기술평의원.

2) Marx, V. (2013), "The Big Challenges of Big Data", Nature, 498: 255-260.

3) 신은정 (2015), 「오픈사이언스(Open Science)에 관한 OECD 논의 동향과 시사점」, 동향과 이슈 제2호, 과학기술정책연구원.

4) 유럽 생물정보학 연구소(EMBL-EBI)의 연구데이터 공유/재활용을 통한 경제적 이득은 연간 1조 파운드, 데이터 및 서비스 연구지원 연간 9.2억 파운드로 추정(Beagrie and Houghton, 2016)

않고 연구를 수행함으로써 새로운 과학적 발견을 가능케 하는 효과도 있다.

특히, 공개적이고 자유롭게 논의하고 다양한 방법으로 재현과 입증을 통해 연구에 대한 정확한 검증으로 연구 결과에 대한 투명성을 확보할 수 있다. 연구의 결과물이 공개되어 연구 성과에 대한 검증과 재현에 누구나 자유롭게 참여할 수 있기 때문에 표절, 데이터 조작과 같은 부정행위를 방지할 수 있는 긍정적인 기능도 가능하다.

## 나. 연구데이터의 정의

연구데이터에 대한 정의는 국가별로 약간은 상이하나, 기본적으로 연구자가 연구개발사업의 과정이나 결과로서 생산 또는 수집되는 데이터를 의미한다. 최근, 우리나라는 「국가연구개발사업의 관리 등에 관한 규정」에 연구데이터에 대한 정의를 법제화하고, 국가연구개발사업의 연구 성과물로서 관리와 활용을 위한 최소한의 근거를 제시하였다. 각 국가별 연구데이터의 정의는 다음과 같다.

〈표 1〉 주요 국가별 연구데이터 정의

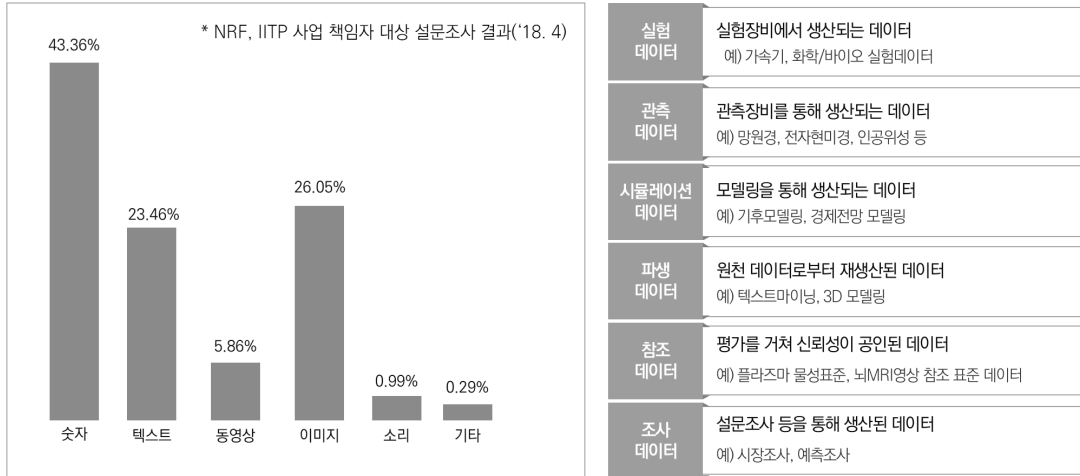
국가	정 의
OECD	과학 연구의 주요 출처로 사용되는 사실적 기록(수치, 문자, 이미지, 음성 등)으로 정의되며, 연구 결과를 검증하는 데 필요한 데이터
미국	연구과정의 전 주기에서 발생하는 데이터이며, 연구자가 수집할 정보와 수집 방법, 정보 처리, 분석 계획 등을 포함
영국	연구데이터 관리에 관한 지침서를 통해 데이터 관리차원에서 연구의 전 주기에서 생성되는 데이터
호주	사실, 관찰, 이미지, 컴퓨터 프로그램 결과, 기록, 경험의 형태에서 생성되는 데이터
한국	연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터

연구데이터의 형태는 텍스트, 숫자, 이미지, 동영상 등 다양하다<sup>5)</sup>. 국가 연구개발사업을 수행한 연구자들을 대상으로 연구데이터 형태를 조사한 결과<sup>6)</sup>를 살펴봐도 유사함을 알 수 있다. 연구데이터의 종류와 유형도 매우 다양하며 생산되는 방식에 따라서 다양한 종류의 데이터가 생산되고 있다.

5) 이상환·심원식 (2009), 「과학데이터의 공유와 활용」, KISTI 지식리포트, 한국과학기술정보연구원.

6) 연구재단(NRF), 정보통신기획평가원(IITP) 사업 책임자 대상으로 설문조사('18. 4)

〈그림 1〉 연구데이터 형태 및 유형



## 2

### 국내외 현황

#### 가. 해외 현황

##### ① 영국

영국은 연구회 중심으로 연구데이터를 관리 및 활용하는 체계로 이루어져 있다. 먼저, 법/제도적인 관점을 살펴보면, 정보자유법(The Freedom of Information Act, 2000)에 의거하여 국가 차원에서 재사용 가능한 형태로 연구데이터 공개를 의무화하고, 데이터 정책에 관한 공동원칙을 기반으로 분야별 연구위원회에서 세부적인 규정 제시하고 있다<sup>7)</sup>. 특히, 2013년에는 데이터셋을 재사용이 가능한 포맷으로 제공하도록 정보자유법 개정하였으며, 제 3 자가 사용할 수 있는 라이선스 허용 항목 명시에 대한 내용을 추가하였다.

7) 심원식 (2015), “국가 차원의 연구데이터 관리체계 구축을 위한 로드맵 제안 - 영국 사례 분석을 중심으로 -”, 「한국문헌정보학회지」, 49(4): 355-378.

〈표 2〉 UKRI의 Common Principles on Data Policy

- ① 공적자금으로 수행된 연구데이터는 최소한의 제한으로 적시에 활용가능
- ② 연구특성을 반영한 데이터 관리정책과 계획 시행 및 장기적 가치를 지닌 데이터의 보존과 접근성 유지
- ③ 타 연구자들이 데이터 활용을 위한 충분한 메타데이터 공개
- ④ 연구데이터 품질을 위해 모든 연구단계에서 관련 정책 고려
- ⑤ 연구자가 일정기간 독점적으로 데이터 사용
- ⑥ 데이터 출처 표기 및 이용조건 준수
- ⑦ 데이터 공유를 위한 공적자금 사용

특히, 데이터 관리 계획(Data Management Plan, 이하 DMP)의 의무제출 규정은 7개 연구위원회 중에서 6개가 시행하고 있으나, EPSRC는 출판물에 연구데이터 정보 및 접근방안을 기재하도록 하고 있다. 연구데이터 공유·활용 추진체계는 분야별 연구위원회가 공공 연구과제에서 생성된 데이터와 성과물을 관리 및 공유하고 있으며, 연구자를 지원하기 위한 디지털 큐레이션 센터(Digital Curation DCC)도 운영하고 있다.

연구데이터 관련 정책 추진은 7개의 분야별 연구위원회가 독립적으로 R&D 연구 수행·지원, 산출물 관리, 공유정책을 수립 및 추진하고(심원식, 2015) 있으며, 영국연구위원회(UKRI)는 분야별 연구위원회의 협의체로 전체 연구위원회의 최고 정책결정기구로서의 역할을 하며 분야별 연구위원회의 공동 활동 촉진 활동을 수행하고 있다.

연구데이터 관리 및 공유 활용 지원을 위해 디지털 큐레이션 센터에서는 DMP 작성 지원, 데이터 관리자 교육 등을 지원하고, 정보시스템합동위원회(JISC)에서는 연구데이터 공유 인프라를 지원하고 있다. 대학 내 연구자들은 데이터의 중요성 및 관리의 필요성을 인식하고 있으며, 주로 내부적으로만 데이터를 공유하며 연구 그룹 규모에 따라 내부 관리 혹은 리포지토리를 구축하여 데이터 관리 및 공유하고 있다.

## ② 미국

미국은 연방기관 중심으로 연구데이터 관리 및 활용하는 체계로 이루어지고 있다. 미국도 영국과 마찬가지로 정보자유법이 공공 연구결과물의 개방과 활용에 대한 법적인 근거이며, 정부가 연구결과물 공유·활용을 위한 지침을 제시하고 연구관리기관들이 세부적인 규정을 수립하고 있다. 특히, 과학기술정책국(OSTP)은 공공 연구결과물의 접근성 향상을 위한 지침(Memorandum, 2013)을 통해 행정부 및 산하 기관을 대상으로 공적 연구결과물의 공유·활용을 위한 원칙 제시<sup>8)</sup>하고 있다.

8) 연간 1억 달러 이상 연구비를 운용하는 연방 기관은 연구 논문 및 데이터에 대한 효율적인 접근 방안 계획(Public Access Plan)제출

연구데이터 공유·활용을 위한 추진체계는 과학기술정책국, 예산정책국 등이 연방 차원의 법제도 및 원칙을 수립하고, 이를 기반으로 연구관리기관들이 세부적인 연구데이터 공유·활용 계획을 수립한다. 연간 1억 달러 이상 연구개발비를 운용하는 국립과학재단(NSF), 국립보건원(NIH), 에너지부(DOE) 등 연방기관은 연구데이터 공유·활용 추진 조직의 역할을 수행하고 있다.

연구데이터 관리 및 활용 지원조직은 주로 대학 도서관이 대학 소속 연구자들의 DMP 제출, 데이터 관리 및 공유를 지원하는 역할을 수행하고 있다. 또한, 주요 연방기관 및 대학에서는 연구데이터 관리 및 공유를 위한 국가 데이터 센터와 리파지토리를 운영하고 있으나 각 시스템을 총괄하거나 연계하는 정책과 체계는 부재하다.

미국의 연구데이터 공유·활용 인식은 다수의 연구자들이 데이터 공유·활용 필요성을 인지하고 있어, 미국 내 연구자 중 46%가 데이터 공유 경험이 있다. 이들 중 66%가 데이터 공유를 연구사회 규범으로 인식하고 있다<sup>9)</sup>.

### ③ 호주

호주는 영국과 마찬가지로 연구회를 중심으로 연구데이터 관리 및 활용하는 체계로 이루어져 있다. 공공 연구과제에서 생산된 연구데이터는 공공 정보로 간주하여 공개와 접근이 법적으로 보장되어 있다. 호주는 국가 지침<sup>10)</sup>을 통해 연구데이터를 포함한 연구 산출물의 관리 방안을 제시하고 개별 연구기관들이 세부 규정<sup>11)</sup>을 수립하고 있다.

연구데이터의 공유·활용 추진체계는 연구관리기관이 제시한 지침을 기반으로 개별 연구 수행기관들이 자체 규정을 수립 및 실행하고 호주대학연합과 호주국가데이터서비스(Australian National Data Service, ANDS)가 연구데이터 관리 및 공유 정책 준수를 위한 지원 활동을 수행하고 있다. 호주연구위원회(ARC)와 보건의료연구위원회(NHMRC)가 연구결과물에 대한 관리와 공유 방향을 제시하는 기본 지침(the Code)을 수립하고, 연구기관 및 대학과 같은 연구 수행기관이 연구데이터 관리 및 공유에 관한 세부 규정을 수립하여 시행하고 있다.

연구데이터 관리 및 활용 지원은 호주국가데이터서비스를 운영하고 호주

---

의무화하였음(김지현, 2013). 해당 연방기관들은 Public Access Plan을 통해 데이터관리계획(Data Management Plan, DMP) 제출을 의무화하고 DMP 타당성 및 이행여부 평가결과를 과제선정 및 연구비 지급 여부에 반영하고 있음.

9) Ferguson, L. (2014), "How and Why Researchers Share Data (and Why They don't)", <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont> (28 Nov 2019).

10) 2007년 발표된 '책임 있는 연구수행을 위한 국가지침(Australian Code for the Responsible Conduct of Research, the Code)'을 통해 연구 데이터를 포함한 연구 산출물의 생산 및 관리 규정 제시

11) 호주연구위원회(ARC)는 2014년부터 ARC의 연구비를 지원 받는 연구기관에 대하여 의무적으로 DMP을 제출할 것을 규정하였으며 DMP 세부 지침 및 템플릿 제공 관련 책임은 개별 연구기관 및 대학에 부여

대학연합은 각 대학의 연구데이터 정책 수립 및 지원을 수행하고 있다. 또한, 국가차원에서 연구데이터의 수집, 공유, 관리 등 전주기에 대한 인프라가 체계적으로 구축되어 있으며 대학에서도 개별적인 플랫폼 운영하고 있다.

연구데이터 관리 및 활용 지원은 호주국가데이터서비스를 운영하고 호주 대학연합은 각 대학의 연구데이터 정책 수립 및 지원을 수행하고 있다. 또한, 국가차원에서 연구데이터의 수집, 공유, 관리 등 전주기에 대한 인프라가 체계적으로 구축되어 있으며 대학에서도 개별적인 플랫폼 운영하고 있다.

## 나. 국내 현황

우리나라는 데이터 중심 R&D 활성화로 연구데이터 활용 수요가 점증함에 따라, 2018 년에 과학기술정보통신부는 국가 연구개발 사업을 통해 축적되는 연구데이터를 국가 차원에서 공유·활용할 수 있는 체계를 구축하고자 '연구데이터 공유·활용 전략(안)'을 발표하였다. 특히, 연구자가 생산한 연구결과의 검증에 필수적인 연구데이터를 체계적으로 관리하기 위하여 연구개발계획서 작성 시 데이터관리계획을 수립하도록 '국가연구개발사업의 관리 등에 관한 규정'을 개정 및 시행하였다<sup>12)</sup>. 또한, DMP 제도에 대한 우선적인 시행을 위해 과학기술정보통신부 산하 기관에서 공고되는 과제 중에서 일부 과제를 DMP 를 시범 적용하였다<sup>13)</sup>.

우리나라 연구자들이 연구데이터에 대한 관리 및 활용 인식을 살펴보면<sup>14)</sup>, 연구데이터를 생산하는 과제의 수에 비해 생산된 연구데이터를 개방과 공유한 경험이 매우 적었다. 더욱이 자신이 보유하고 있는 연구데이터를 공유하는 것에 비해 다른 연구자들이 생산된 연구데이터를 활용하고 싶다는 생각이 있음을 알 수 있었다. 또한, 많은 연구자들이 연구데이터를 공개하지 않거나 존재와 보유 위치를 몰라서 활용하기 어렵다는 애로 사항을 말하였다.

12) 연구데이터 및 DMP 정의, DMP 제출 의무사항, DMP 보고 및 DMP 관리 및 연계 등을 공동관리규정에 반영하여 개정('19. 3. 19)하고 시행하였음('19. 9. 1)

13) 연구재단 4개 분야 10개 사업과 정보통신기술원은 4개 분야 10개 과제에 DMP 시범 적용하였음('19. 2)

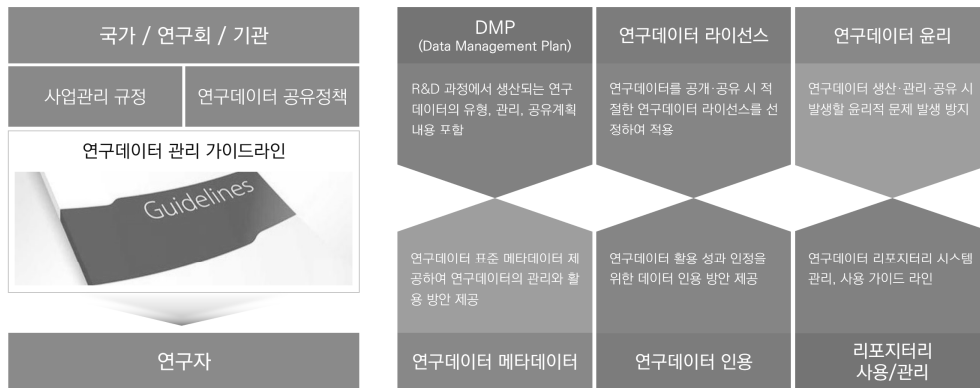
14) 국가과학기술연구회 산하 출연(연) 연구데이터 관리 및 공유 실태 조사 실시(KISTI, 2018)

### 가. 거버넌스 체계 수립

현대 과학의 연구 방식이 데이터 중심으로 연구 패러다임이 전환되고 공유와 협력에 기반하는 오픈 사이언스(Open Science) 정책 대두되고 있다. OECD, 미국, 영국, 호주 등 주요 선진국에서는 연구데이터를 체계적인 관리, 쉬운 접근, 생산, 보존 및 데이터의 재사용을 통한 가치 창출노력 수행하고 있다. 따라서, 기관 또는 국가 관점에서도 연구데이터의 관리 및 활용을 위한 거버넌스 체계를 고려해야 한다.

연구 분야의 특성을 반영하여 연구데이터 공유 정책, 데이터 관리 계획(Data Management Plan) 등이 포함된 세부적인 연구데이터 공유 및 활용 정책을 수립하고 추진해야 한다. 가이드라인, 프로토콜, 표준 지침 등을 포함하는 국가 또는 기관 차원의 정책과 이행 절차는 연구데이터 관리를 위해 필수적이다.

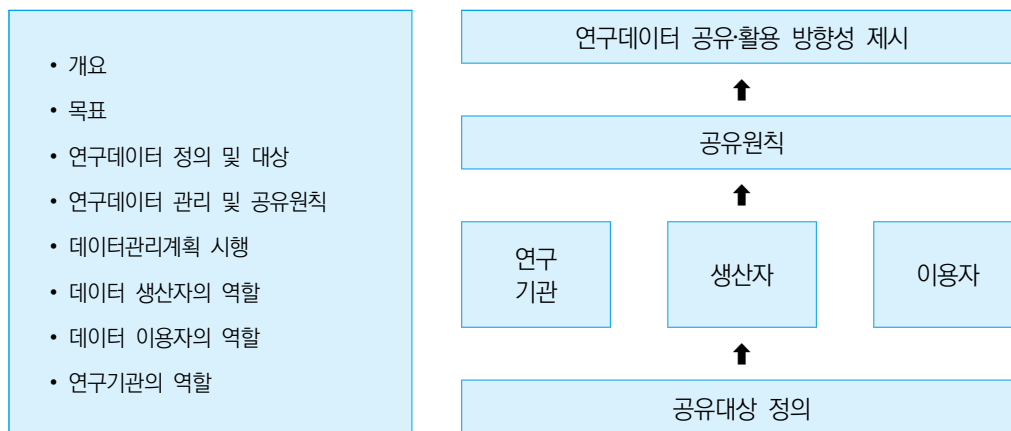
〈그림 2〉 연구데이터 관리 및 활용을 위한 거버넌스 체계



### 나. 연구데이터 공유 정책

연구데이터 공유와 활용의 목표를 달성하기 위해서는 기관 및 연구자들에게 연구, 교육 및 공익을 위한 목적으로 더욱 광범위한 데이터 이용을 위해 장기적 관점에서 연구데이터의 가용성을 보장해야 한다.

〈표 3〉 연구데이터 공유 원칙 구성 요소



연구개발사업을 통해 생산된 연구데이터는 중요한 기관 또는 국가의 자산으로 관리·보존하여 사회적 공공재로서 누구나 자유롭게 사용할 수 있도록 공유되어야 한다. 이러한 연구데이터에 대한 개방적인 접근은 국가 연구개발의 혁신을 촉진하고 고품질 연구를 가능하게 한다. 해외에서는 공적 자금을 통해 생산된 연구데이터는 공개가 기본 원칙이나, 보안과제, 개인정보 보호 및 계약에 따른 공개금지 조건이 포함된 연구, 데이터에 대해서는 접근을 제한할 수 있다.

#### 다. 데이터관리계획(Data Management Plan)

기관 또는 국가에서는 연구개발 과제 기획 시 연구 결과로 생성되는 연구데이터의 관리, 공유 및 활용을 위한 데이터관리계획을 작성하는 가이드라인을 제공해야 한다. 일반적으로 R&D 과제를 통해 생산되는 데이터의 개요, 관리, 공유 계획을 담고 있는 1~2 쪽 분량의 문서이다. DMP 제도의 중요한 점은 연구자들에게 데이터의 체계적인 관리와 공유 의무에 대한 인식을 고취시키는 중요한 계기가 될 수 있다. 또한, 국가 또는 기관 차원에서 수행된 연구개발 과제로부터 생산되는 데이터 현황을 파악할 수 있다.

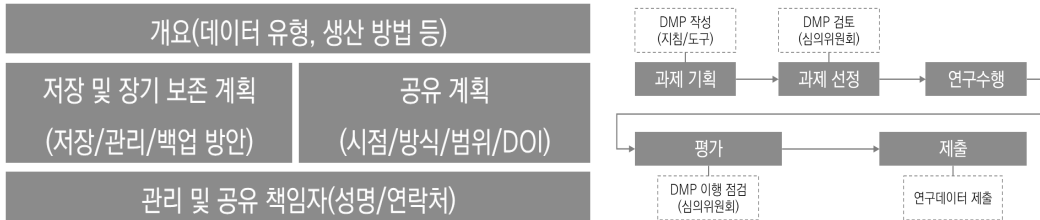
데이터관리계획 검토는 연구계획의 합리성, 적합성 및 광범위한 영향 모두를 평가하기 위해 중요한 사항이다. 연구계획서는 때때로 많은 커뮤니티 연구원과의 협력을 필요로 하므로, 제출된 데이터관리계획은 해당 연구 분야의 실무 커뮤니티가 개발한 절차, 표준 및 모범 사례를 고려해야 한다. 특히, 연구자들을 대상으로 데이터관리계획 작성에 대한 직접적인 지원, 질문에 대한 답변 및 컨설팅 등 실질적으로 지원해야 한다.

DMP 는 국가별, 분야별 다양한 형식이 존재한다. 그러나, DMP 최소한의



형식 요건은 연구과제에서 생산되는 데이터와 관련하여 연구과제명, 연구 데이터 개요, 연구데이터 보존계획, 연구데이터 공유계획, 데이터 관리 및 공유 책임자로 구성한다.

〈그림 3〉 DMP 구성 요소 및 프로세스 절차



## 라. 연구데이터 라이선스

연구데이터는 일부 공유 제한사항을 제외하고 다른 연구자들이 재사용할 수 있도록 공개하는 것이 권장되나, 가능한 이용허락조건을 명확하게 제시해야 한다. 특히, 연구데이터를 상업적으로 이용하게 하거나 유료 제공 또는 데이터 제공에 비용이 드는 경우 그 조건을 명시해야 한다. 라이선스는 연구데이터에 대한 저작자의 권리를 보호하고 데이터의 재사용을 촉진하기 위한 이용허락조건을 설정하거나, 데이터 제공자와 이용자의 권한과 의무를 명시할 수 있는 연구데이터 라이선스와 이용 표준협약을 제시할 수 있다.

## 마. 연구데이터 표준 메타데이터

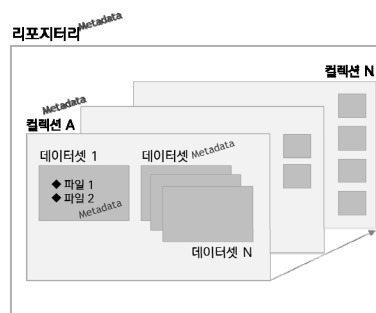
일반적으로 해외에서는 연구데이터를 발견하기 쉽고, 재사용 가능하게 만들기 위한 FAIR 원칙<sup>15)</sup>을 준수하고 있다. 연구 분야나 주제의 특성에 따라 메타데이터가 매우 다양할 수 있다. 연구데이터 리포지터리 관점에서 관리 및 활용하기 위한 메타데이터 구조는 <표 4>와 같이 제안한다<sup>16)</sup>.

15) <https://www.force11.org/group/fairgroup/fairprinciples>

16) 연구데이터 관리 및 공유를 위한 메타데이터(2017, TTA 표준)

〈표 4〉 연구데이터 메타데이터 구성 요소

관리 및 공유, 재사용의 가치가 있는 개별 단위의 연구데이터
개별 파일에 대한 메타데이터의 그룹으로 구성(파일 없이 메타데이터 단독으로 구성 가능)
데이터 셋의 논리적 그룹으로, 그룹 단위는 프로젝트, 부서, 연구과제 등 자유롭게 설정
기관에서 연구데이터를 수집, 저장, 관리, 출판하기 위한 시스템 의미



## 바. 연구데이터 인용

CODATA-ICSTI Task Group 에서 데이터 인용은 학술적 크레딧을 부여 하고 데이터에 대한 접근을 용이하도록 데이터에 대한 참조로 정의하고 있다. 연구에서 이용되는 데이터 셋이 크고 복잡해지면서 논문의 일부분으로 데이터를 출판하는 것이 아니라, 연구 결과에 대한 과학적 주장을 검증하기 위해 출판물과 해당 데이터와의 연계 요구가 커지고 있다. 데이터 생산자에 대한 크레딧으로써 데이터 인용과 활용도 측정이 필수적이나, 연구데이터의 활용도를 측정하기 위해서는 데이터 인용이 선행되어야 한다. 2013 년 Beyond the PDF 2 Conference 에서 8 가지 원칙을 제시하였고, 2014 년 Force11 에서 발표한 데이터인용 원칙에 관한 통합 선언(Joint Declaration of Data Citation Principles)<sup>17)</sup>이 가장 최신의 원칙이다.

〈표 5〉 Force11 Joint Declaration of Data Citation Principles

원칙	정의
중요도 (importance)	<ul style="list-style-type: none"> <li>• 데이터는 합법적이고 인용 가능한 연구 결과물임</li> <li>• 데이터인용은 출판물과 동일한 중요도가 부여되어야 함</li> </ul>
기여 인정과 기여자 표시 (credit and attribution)	<ul style="list-style-type: none"> <li>• 데이터인용은 데이터의 모든 기여자에게 학술적인 인정과 규범적이고 법적인 기여자 표시를 제공해야 함</li> <li>• 단일의 기여자 표시 방식이 모든 데이터에 적용 가능하지는 않다는 점도 인지해야 함</li> </ul>
증거 (evidence)	<ul style="list-style-type: none"> <li>• 학술 문헌이나 다른 어떤 경우에도 데이터가 사용될 때 상응한 데이터가 인용되어야 함</li> </ul>
고유 식별 (unique identification)	<ul style="list-style-type: none"> <li>• 데이터인용은 기계 작동이 가능하고, 세계적으로 고유하며, 커뮤니티에서 폭넓게 사용되는 지속적인 방식을 포함해야 함</li> </ul>
접근 (access)	<ul style="list-style-type: none"> <li>• 데이터인용은 데이터 자체에 대한 접근뿐만 아니라 관련된 메타데이터, 문서, 코드, 다른 자료 등이 해당 데이터에 대한 이용 정보로써 기계와 인간에 모두 이용 가능해야 함</li> </ul>

17) <https://www.force11.org/datacitationprinciples>

원칙	정의
지속성 (persistence)	• 고유 식별자와 데이터를 기술한 메타데이터와 배치는 지속적이어야 하고, 심지어 데이터의 수명보다 길어야 함
구체성과 검증가능성 (specificity and verification)	• 데이터인용은 구체적인 데이터의 식별, 접근, 검증가능성을 제공해야 함 • 인용과 인용 메타데이터는 데이터의 특정 시분할, 버전, 세분된 영역을 검증하기에 충분할 정도로 출처와 무결성에 대한 정보를 제공해야 함
상호운용성과 유연성 (interoperability and flexibility)	• 데이터인용은 커뮤니티 간 상호운용성을 저해하지 않는 수준에서 여러 커뮤니티의 다양한 데이터 인용 방식을 충분히 포괄할 수 있을 정도로 유연해야 함

연구데이터의 인용 포맷은 해외 출판사나 DataCite에서도 제공하고 있다. 우리가 제안하는 연구데이터의 인용 포맷<sup>18)</sup>을 사용하여 논문 등 출판물의 참고문헌 섹션에서 인용할 것을 권장한다.

<표 6> 연구데이터 인용 포맷

Creator (Publication date): Title. Version. Location of publisher: Publisher. (Resource type). Identifier. Access date.
-------------------------------------------------------------------------------------------------------------------------

데이터 인용 요소의 세부적인 설명은 <표 7>과 같다.

<표 7> 데이터인용 요소

요소	정의
Creator (생산자)	• 데이터 생산에 참여한 주요 연구자 (기여도 순)
Publication date (출판일)	• 다음 중 최신의 날짜 : 데이터셋이 이용한 날짜, 품질 보증 절차가 완료된 날짜, 엠바고 기간이 만료된 날짜
Title (제목)	• 데이터셋의 명칭
Version (버전)	• 데이터셋의 버전 (버전은 데이터가 추가되거나 변경될 때 증가함)
Resource type (자원 유형)	• 데이터셋 유형을 기술
Publisher (출판자)	• 데이터셋을 보관, 저장, 배포하는 리포지터리, 시스템, 조직 등의 이름
Identifier (식별자)	• 통용되는 영구적 식별체계를 따르는 데이터 식별자
Access Date (엑세스 날짜)	• 해당 데이터셋에 접근한 최근 날짜
Location of publisher (출판자 위치)	• 출판자의 위치

18) 정은경 외 (2018), 「연구데이터 활용지표 개발 연구」, 한국과학기술정보연구원(재인용).

## 4

## 결론

첫째, 연구데이터에 대한 연구자의 인식 전환이 필요하다. 우리나라의 연구자들이 연구데이터에 대한 인식은 기본적으로 연구 활동을 통해 생산된 연구데이터는 연구자 개인 소유로 인식하고 있다. 그러나, 해외 주요국처럼 국가의 공적 자금을 통해 생산된 연구데이터는 국가의 소유, 즉 공공재로 인식하고 연구데이터를 개방과 공유를 통해 데이터 활용의 가치를 높이고 있다. 이러한 연구데이터 개방과 공유는 불필요한 연구의 중복을 최소화함으로써 연구개발의 효율성을 증대할 수 있기 때문이다.

둘째, 연구데이터의 공유는 새로운 가치를 창출한다. 일반적으로 많은 연구자들이 인식하는 것처럼 연구데이터의 공유가 손해가 아니라, 새로운 과학적 발견을 위한 근간이다. 희스 입자 발견 등과 같이 대용량의 연구데이터를 다수의 연구자들이 함께 공유 분석함으로써 기존의 연구 가설을 검증한다. 소재물성 데이터, 코호트 데이터 등과 같은 연구데이터를 이용하여 신소재 개발, 신약 개발, 질병 치료 등 새로운 부가가치를 만들 수 있기 때문이다.

셋째, 데이터관리계획 제도의 추진으로 연구데이터 공유 문화를 정착할 필요가 있다. 연구자의 연구 환경이 디지털 환경으로 전환됨에 따라 데이터 기반으로 연구를 수행하여 많은 연구데이터를 생산 분석하고 있다. 그러나, 많은 연구자들이 누가 생산하였지, 어느 곳에 저장되었는지, 공개 및 활용 가능인지에 대한 정보가 없어서 기존의 연구데이터를 활용하기 힘들다고 어려움을 토로하고 있다.

넷째, 활용 중심의 연구데이터 거버넌스 체계 구축해야 한다. ICT의 발전과 측정 및 관측 장비의 발달로 인해 많은 연구데이터가 생산됨에 따라, 생산되는 연구데이터 양에 비해 연구데이터 수집 및 관리할 수 있는 기술과 인프라의 한계가 존재한다. 또한, 디지털 자원의 특성 중 하나인 휘발성을 고려하여, 연구데이터를 활용 시 장기적인 보관과 폐기 정책도 고려할 필요가 있다. 데이터 생태계는 다양한 이해 관계자(과제관리자, 연구자, 이용자 등)들이 존재하며, 각 이해 관계자들이 데이터를 잘 공유하고 활용할 수 있는 생태계를 구축할 필요가 있다.