OPEN ACCESS

과학기술분야 정부출연연구기관 연구데이터 관리 방안 연구*

A Study on Research Data Management Methods for Government-funded Research Institutes in the Field of Science and Technology

한 나 은 (Na-eun Han)**
엄 정 호 (Jung-Ho Um)***
임 형 준 (Hyung-Jun Yim)****

목 차

1. 서 론
2. 이론적 배경
3. 연구데이터 관리 현황
4. 연구데이터 공유 체계 구축 방안
5. 결론 및 제언

초 록

본 연구는 연구데이터의 공유 및 활용 촉진을 목적으로 NST 산하 출연(연)의 연구데이터 관리 현황을 분석하고, 이를 바탕으로 연구데이터 공유 및 관리 체계 구축을 위한 방법을 제언한다. 연구데이터 관리 현황 조사는 총 20개의 출연(연)을 대상으로 22년도 23년도, 두 차례에 걸쳐 진행되었으며, 기관별 보유 연구데이터 현황, 기관별 연구데이터 관리 정책 현황, 기관별 연구데이터 관리시스템 현황을 분석하였다. 뿐만 아니라 연구데이터의 관리 및 공유에 있어서 어려움과 개선되어야 할 점을 파악하고, 이를 기반으로 연구데이터 공유 및 관리 체계 구축을 위한 방법을 정책적 측면, 시스템적 측면, 연계 체계 구축 측면으로 나누어 제언하였다. 연구데이터 공유 체계 구축을 위해서는 정책적 근거를 마련하여 연구데이터의 정의, 적용 범위, 관리의 내용, 활용 방식, 선도기관 등의 내용을 제시하는 것이 바람직할 것이다. 또한 체계적이고 일원화된 연구데이터 관리를 위해 기관별로 리포지토리 및 관리시스템을 구축하여 관리하되 이를 DataON과 연계하고 원스톱 서비스를 제공함으로써 데이터의 접근성 및 활용도를 제고할 수 있을 것이다.

ABSTRACT

This study analyzes the current status of research data management at NST-affiliated government-funded research institutes for the purpose of promoting the sharing and use of research data, and based on this, suggests methods for establishing a research data sharing and management system. The survey on the status of research data management was conducted twice in 2022 and 2023 for a total of 20 research institutes. In addition, difficulties and areas that need to be improved in the management and sharing of research data were identified, and based on this, methods for establishing a research data sharing and management system were proposed by dividing them into policy aspects, system aspects, and linkage system construction aspects. In order to establish a research data sharing system, it would be desirable to prepare a policy basis and present contents such as the definition of research data, scope of application, contents of management, utilization method, and leading institutes. In addition, for systematic and unified research data management, it would be recommended that each institute will establish and manage a repository and management system. By linking this with DataON, the national research data platform, and providing one-stop services, the accessibility and usability of data will be improved.

키워드: 연구데이터, 연구데이터 관리, 국가연구데이터플랫폼, 연구데이터 리포지토리, 연구데이터 공유
Research Data, Research Data Management, National Research Data Platform(DataON), Research Data Repository, Research Data Sharing

* 본 연구는 한국과학기술정보연구원(KISTI) 연구사업(과제번호: K24L1M2C3)의 지원에 의해 이루어진 것임.
** 한국과학기술정보연구원(KISTI) 박사후연구원(betterhan@kisti.re.kr / ISNI 0000 0004 9212 2493) (제1저자)
*** 한국과학기술정보연구원(KISTI) 책임연구원(jhum@kisti.re.kr / ISNI 0000 0004 7399 3760) (공동저자)
**** 한국과학기술정보연구원(KISTI) 선임연구원(hjyim@kisti.re.kr / ISNI 0000 0004 7393 6934) (교신저자)
논문접수일자: 2024년 4월 22일 최초심사일자: 2024년 4월 30일 게재확정일자: 2024년 5월 15일
한국문헌정보학회지, 58(2): 151-175, 2024. http://dx.doi.org/10.4275/KSLIS.2024.58.2.151

※ Copyright © 2024 Korean Society for Library and Information Science
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서 론

1.1 연구의 배경 및 목적

연구데이터란 일반적으로 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰조사 및 분석 등을 통하여 산출된 사실자료로서 연구결과의 검증에 필수적인 데이터”를 의미하며(과학기술정보통신부, 2020), 이는 연구성과물인 연구보고서나 논문과는 구별되는 개념이다(한나은, 김성희, 2014). 연구데이터는 연구결과에 대한 핵심정보를 담고 있기 때문에 연구결과를 검증 및 증명하기 위해 필수적인 자료라고 할 수 있으며(한나은, 2022b; Pilat & Fukasaku, 2007), 연구데이터 공개를 통해 중복 연구 및 데이터의 중복 생산을 방지함으로써 연구성과를 제고하고 연구 효율성 향상을 기대할 수 있다(한국과학기술정보연구원, 2019). 연구데이터를 생산하는 다양한 주체가 존재하는데, 그 중 주요한 집단 중 하나가 국가과학기술연구회(National Research Council of Science & Technology, 이하 NST) 산하 정부출연연구기관(이하 출연(연))이라고 할 수 있다. 출연(연)에서 생산하는 연구데이터는 국가 지원을 받아 진행되는 연구 과정에서 발생한다는 측면에서 개별 연구자들이 생산하는 연구데이터와는 차별적인 특성을 갖는다. 국내 총 25개의 출연(연)에서 지속적으로 생산하는 연구데이터의 양이 상당함에도 불구하고 출연(연) 연구데이터의 관리 및 현황 분석에 관한 연구는 활발하지 못한 편이며, 최근 출연(연) 연구데이터와 관련해서는 인프라 및 서비스 측면에 초점을 맞춘 연구가 진행된 바 있다(박성은 외, 2024).

이미 전 세계적인 연구의 흐름은 제4의 과학 패러다임이라고 불리는 데이터 집중형 과학, 데이터 집중형 연구로 전환되었으며, 이에 따라 연구데이터의 공유, 관리 및 재활용의 중요성은 지속적으로 강조되고 있다(국가과학기술연구회, 2019; Tolle, Tansely, & Hey, 2011). 연구데이터의 적극적인 공개 및 공유가 지속되기 위해서는 체계적인 관리가 필수적인데, 이러한 체계적인 관리가 시행되기 위해서는 경제적, 정책적, 인적 문제 등 다양한 문제들을 고려해야 할 필요가 있다. 연구데이터의 공유 및 재활용의 중요성은 이미 다양한 연구에서 진행된 반면(김예현, 김지현, 2023; 김주섭, 김선태, 2023; 한나은, 2022a), 이를 체계적으로 수행하기 위한 다양한 접근에 대한 연구를 출연(연) 연구데이터를 중심으로 살펴본 연구는 거의 없는 실정이다. 본 연구는 국내 출연(연) 연구데이터 관리 현황을 분석함으로써 현재 생산되고 있는 연구데이터가 어떻게 관리 및 공유되고 있는지, 관리 및 공유에 있어서 어려움과 개선되어야 할 점은 무엇인지를 파악하고, 이를 기반으로 연구데이터 공유 및 관리 체계 구축을 위한 방법을 정책적 측면, 시스템적 측면, 연계 체제 구축 측면으로 나누어 제언한다는데 의의를 갖는다.

1.2 연구의 내용 및 방법

본 연구는 연구데이터의 공유 및 활용 촉진을 목적으로 NST 산하 출연(연)의 연구데이터 관리 현황을 분석하고, 이를 바탕으로 연구데이터 공유 및 관리 체계 구축을 위한 방법을 제언한다. 출연(연) 연구데이터 관리 현황 파악을 위

해 해당 연구기관의 연구데이터 담당자를 대상으로 온라인을 통한 설문을 두 차례 실시하였으며, NST 산하 출연(연) 가운데 설문 수행 주체인 한국과학기술정보연구원(Korea Institute of Science and Technology Information, 이하 KISTI)과 국가보안기술연구소(National Security Research Institute, 이하 NSR)를 제외한 총 23개의 출연(연)에 설문을 송부하였다.

1차 설문조사는 2022년 9월, 약 한 달 동안 서면을 통해 실시하였으며, 설문 문항은 기관의 보유 연구데이터 및 관리 시스템 현황, 관리 정책 및 전담 관리 인력 현황, 공개 가능 데이터 및 연구데이터 공유 현황 등에 대한 질문 내용을 포함한다. 총 23개 출연(연) 가운데 20개 기관이 회신하였으며 응답률은 86.9%이다.

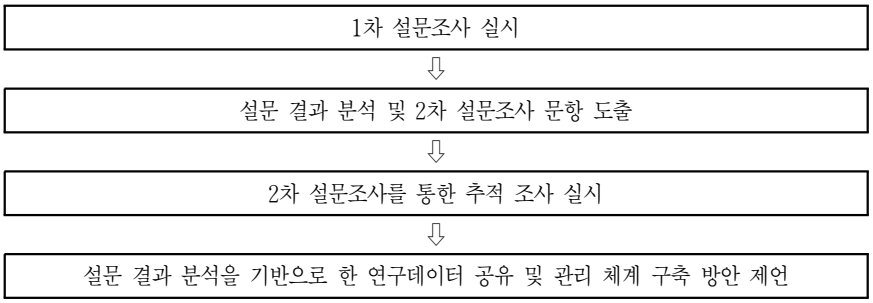
2차 설문조사는 2023년도 12월에 실시되었으며, 기관의 보유 연구데이터 및 공개 가능 데이터, 관리 시스템 현황에 초점을 맞추어 추적 조사를 수행하기 위해 2022년도에 실시한 설문 문항 가운데 일부 문항을 중심으로 설문을 재실시하였다. 응답률은 약 70%로 총 23개의 출연(연) 가운데 16개의 기관이 2차 설문조사에 회신하였는데, 이는 1차 설문조사에 응답한 기관 20개 가운데 16개의 기관이 재응답한 것으로써 재응답률은 80%이다. 설문조사에 회신한 20개의 기관명은 익명을 보장하기 위해 ID로 대체하였다.

본 연구는 두 차례에 걸쳐 진행한 설문 내용을 바탕으로 기관별 보유 연구데이터 현황, 기관별 연구데이터 관리 정책 현황, 기관별 연구데이터 관리시스템 현황의 내용을 분석하여 제공하고, 연구데이터 공유 및 관리 체계 구축을 위한 방안을 제언한다. 설문을 포함한 연구 절차를 도식화하면 다음 〈그림 1〉과 같다.

2. 이론적 배경

2.1 연구데이터 관리 및 연구데이터 플랫폼

연구데이터가 적극적으로 공개 및 공유되었을 때 예상되는 긍정적인 효과를 바탕으로 연구데이터의 공유, 관리 및 재활용의 중요성이 증대되고 있다. 특히 국내・외의 보조금 지원기관 및 출판사 등은 연구데이터의 체계적인 관리를 보다 적극적으로 요청하고 있는데, 미국국립과학재단(National Science Foundation, 이하 NSF)에서는 NSF로부터 보조금 지원을



〈그림 1〉 연구 절차 도식화

받는 연구를 대상으로 데이터 관리 계획(Data Management Plan, 이하 DMP)을 필수적으로 제출해야 함을 명시하고 있으며, 미국국립보건원(National Institutes of Health, 이하 NIH) 역시 DMP 작성을 요구하고 있다(National Science Foundation, 2014). 국내에서는 2019년, 국가연구개발사업의 관리 등에 관한 규정을 개정 및 시행하여 과학기술정보통신부 산하 기관에서 공고하는 일부 과제에 대하여 DMP를 작성 및 제출하도록 요청하고 있다(과학기술기본법, 2020).

정보통신기술의 발달로 대용량의 연구데이터가 폭발적으로 생산 및 활용되는 제4의 과학 패러다임이라고 불리우는 4세대 연구 패러다임이 등장하였는데, 4세대 연구 패러다임은 데이터 분석을 통해 새로운 이론이나 현상을 발견하는 데이터 중심의 연구라고 할 수 있다. 데이터 중심의 연구가 활성화 됨에 따라 연구데이터의 생산량은 급격하게 증가하고 있으며, 전 세계의 연구데이터 리포지토리 정보를 제공하는 re3data.org에 등록된 리포지토리의 수는 2024년 5월 기준으로 3,207개에 달하고 있다. 이는 2020년 12월 약 2,500개에 달하던 것에 비교할 때 약 130%가까이 증가한 것으로 이해할 수 있다. 뿐만 아니라 전 세계적으로 새로운 지식 가치의 창출과 공유 및 확산을 위해 공적 자금이 투입된 연구성과 및 과정을 개방화하는 오픈 사이언스(Open Science) 정책이 적극 추진되고 있다. 미국은 2013년에 이미 연구데이터 관리 및 공유에 대한 지침을 마련하였고, 유럽연합은 2015년부터 OpenAIRE2020 Project를 시작한 바 있다. 또한 여러 나라에서 연구데이터 관리의 제도화, 연구데이터 공유 및 활용 인프라 지원, 연구자 교육 제공 등을 주요 정책으로 추진 및 지원하고 있다. 해외 주요국의 경우를 살펴보면, 미국은 과학분야별 지역적 허브 구축, 영국은 연구데이터 공유센터 구축, 호주에서는 국가데이터서비스(Australian Research Data Commons, 이하 ARDC)를 구축 및 운영하고 있다.

국내에서는 정부부처에서 '국가 연구데이터 공유·활용' 전략을 통한 국가 차원의 연구데이터의 공유 및 활용을 활성화하려는 노력을 지속하고 있으며, 이와 같은 노력의 일환으로 산재된 연구데이터를 체계적으로 관리·운영하고, 공유하기 위한 국가연구데이터플랫폼(이하 DataON)을 구축하여 매년 고도화를 진행하고 있다(황미녕 외, 2022). DataON은 국내·외 다양한 리포지토리 및 기관에 산재되어 있는 연구데이터를 원스톱으로 검색하고 활용할 수 있도록 접근점을 제공하는 국가연구데이터플랫폼을 지향하며, 각종 연구데이터에 대한 색인, 검색 기능과 연구자가 데이터 융합 및 분석에 활용할 수 있는 분석도구 및 가상작업환경을 구축하여 제공하고 있다.

2.2 연구데이터 공유에 관한 인식

연구데이터 공유에 관한 연구자들의 인식은 이미 다양한 선행 연구를 통해 조사된 바 있다. 연구데이터의 중요성에 관한 인식과 관련하여 2019년 KISTI에서는 국가R&D사업 참여 연구자 2,302명을 대상으로 연구데이터 생산, 관리, 공유, 활용 현황 조사를 실시하였는데, 해당 연구 결과에 따르면 응답자 중 80.7%가 연구데이터 공유를 통해 연구성과를 제공하고 연구 효

율성 향상을 기대한다고 응답했음을 밝혔으며, 동일한 연구에서 국가과학기술지식정보서비스(National Science & Technology Information Service, 이하 NTIS) 등록 연구자 3,246명을 대상으로 한 연구데이터의 경제적 가치에 대한 질문에 있어서는 응답자 가운데 46.9%가 연구데이터의 공유 및 활용이 가능할 경우에 초기 연구 활동을 수행하는데 보다 효율적일 것이라고 답하였다(한국과학기술정보연구원, 2019).

연구데이터의 재활용에 관한 인식에서는 국가R&D사업 참여 연구자 2,302명 가운데 60.9%가 다른 연구자의 연구데이터가 필요한 경우가 많다고 응답한 바 있으며(한국과학기술정보연구원, 2019), 국외에서 총 2,184명의 연구자들을 대상으로 2017년과 2018년에 걸쳐 실시한 설문조사에서는 응답자의 87%에 다다르는 연구자들이 다른 연구자들의 데이터에 접근이 용이하다면 사용할 의향이 있다고 응답하였다. 또한, 동일한 연구의 설문조사에서 본인이 직접 수집한 데이터 외에 다른 연구자들의 데이터를 활용하여 연구를 진행하는 경우의 빈도가 보통 이상이라는 응답 역시 38.3%에 달했다(Tenopir et al., 2020).

동시에 연구데이터 활용의 어려움도 나타났는데, 동일 국외 연구에서는 총 2,184명의 연구자 중 74.6%가 다른 연구자들 혹은 기관의 연구데이터에 접근하는 것에 대한 어려움으로 인해 연구자 본인의 연구를 제한하는 상황이 초래된다고 응답하였으며, 국내에서는 국가R&D사업 참여 연구자 2,302명 가운데 61.4%가 본인이 필요한 다른 연구자의 연구데이터를 확보하는데 어려움이 있다고 답변하였다(한국과학기술정보연구원, 2019). 또한 연구데이터를 수집하는 방법에 있어서 김지현(2012; 2015)은 대학 내의 연구자들을 대상으로 한 두 차례의 설문조사와 인터뷰를 실시하였는데, 설문과 인터뷰 모두에서 연구자들은 다른 연구자들의 연구데이터를 얻는 주요한 방식으로 논문에서 데이터를 직접 추출하거나, 개인적으로 연구자에게 연락하여 연구데이터를 수집하는 경우가 가장 많음을 밝힌 바 있다.

연구데이터 공유에 관한 인식을 조사한 선행연구 결과를 통해 연구데이터의 공유 및 관리의 중요성에 대해서는 많은 연구자들이 공감하고 있는 부분이지만, 현실적으로 연구데이터의 적극적인 재활용을 위한 공유 체계가 구축되지는 않았음을 파악할 수 있다.

3. 연구데이터 관리 현황

3.1 기관별 보유 연구데이터 현황

본 연구에서 실시한 설문에 응답한 20개의 출연(연)이 보유하고 있는 연구데이터의 수는 다음 〈표 1〉과 같다.

2022년과 2023년 모두 설문에 응답한 기관의 경우 연구데이터의 수가 상이한 기관도 있지만, 동일하다고 응답한 기관도 9개 출연(연)에 이른다. 응답 결과에 따르면 보유 연구데이터는 보유하고 있지 않다는 응답부터 260건이 넘는 연구데이터를 보유하고 있다는 응답까지 다양한 편차를 보이고 있음을 알 수 있다. 기관의 특성에 따라 보유 연구데이터의 수에는 상당한 차이가 존재하는데, 이는 실제적으로 연구데이터 수의 차이가 존재하는 것에 더해 연

〈표 1〉 년도별 설문 응답 기관 및 기관별 보유 연구데이터 수

기관 ID	'22년도		'23년도	
	응답 여부	보유 연구데이터 수	응답 여부	보유 연구데이터 수
I-1	○	1건	×	미응답
I-2	×	미응답	×	미응답
I-3	○	0건	○	15건
I-4	○	3건	○	3건
I-5	○	1건	○	1건
I-6	×	미응답	×	미응답
I-7	○	0건	○	62건
I-8	○	2건	○	2건
I-9	○	18건	○	24건
I-10	×	미응답	×	미응답
I-11	○	0건	○	0건
I-12	○	0건	○	261건
I-13	○	1건	○	1건
I-14	○	0건	○	1건
I-15	×	미응답	×	미응답
I-16	○	9건	○	9건
I-17	○	23건	○	23건
I-18	×	미응답	×	미응답
I-19	○	9건	○	10건
I-20	○	0건	○	0건
I-21	○	0건	×	미응답
I-22	○	2건	×	미응답
I-23	○	0건	○	0건
I-24	○	6건	×	미응답
I-25	○	5건	○	2건
응답률	80% (20/25)	-	64% (16/25)	-

구데이터의 수를 파악하는 방법에 대한 차이 때문에 기관별로 차이가 발생하기도 함을 유추할 수 있다. 실제로 2022년과 2023년 모두 설문에 응답한 기관의 경우에 만 1년이 되지 않는 사이에 보유 연구데이터에 대한 편차가 큰 응답이 존재하는 경우들이 있는데, I-7 기관의 경우 2022년도 보유 연구데이터는 0건이라고 응답하였으나 2023년도에는 62건의 연구데이터를 보유하고 있다고 응답하였으며, I-12 기관 역시 2022년도 보유 연구데이터는 0건이라고 응답하였으나, 2023년도에는 보유 연구데이터가 261건이라고 응답하였다. 이는 실제적으로 연구데이터의 생산량이나 보유량이 급격하게 증대했다기보다는 이미 보유하고 있던 연구데이터에 대한 파악이 뒤늦게 이루어진 것이라고 해석하는 것이 더 적절할 것이다. 연구데이터에 대한 일반적인 정의가 존재하지만, 아직까지도 연구자별로 연구데이터를 이해하는 수준에 차이가 있으며, 기관별로도 연구데이터를 계산하는 기준이나, 기관별 관리 주체가 개인 연구자인지 부

서인지 등에 따라 기관 내 보유 연구데이터에 대한 정확한 파악이 어려운 경우가 발생한다.

이에 대한 내용을 보강하기 위해 각 기관별로 기본 사업 가운데 연구데이터가 생산되는 사업의 개수를 파악하였는데, 이는 연구데이터 발생 기본 사업의 개수를 통해 대략적으로 생산 가능한 연구데이터의 양을 유추하여 보유 연구데이터를 파악하도록 함이다. 본 연구에서는 연구데이터가 생산되는 사업의 기준을 중과제와 세부과제로 잡고 설문을 진행하였는데, 중과제와 세부과제의 의미는 다음과 같이 이해할 수 있다. 출연(연)의 경우, 정부에서 지원받은 출연금 사업을 크게 대과제와 세부과제로 분류하는데, 대과제의 예산이 큰 경우에는 해당 과제를 중과제로 다시 나누어 관리할 수 있으며, 이는 각 기관의 운영 규정에 따르게 된다. 이를 기준으로 각 출연(연) 별로 기본 사업 중 연구데이터가 생산되는 사업의 개수와 이 중 연구데이터에 대한 관리가 필요한 과제의 수는 다음 〈표 2〉와 같다.

각 기관별로 연구데이터가 생산되는 사업의 개수의 편차는 존재하지만 평균 6~15개의 중과제를 수행하고 있음에도 불구하고, 앞서 확인한 기관별 보유 연구데이터의 수는 과제의 수에 미치지 못하는 상황이다. 이를 바탕으로 각 출연(연)에서 생산하고 있는 연구데이터를 각 기관에서 정확하게 파악하지 못하고 있음을 유추할 수 있다. 또한 본 연구에서는 설문을 통

〈표 2〉 기관별 연구데이터 생산 사업 개수 및 관리가 필요한 과제의 수

기관 ID	연구데이터 생산 사업 개수	연구데이터 관리가 필요한 과제의 수	비율(중과제 기준)
I-1	중과제 361개	중과제 361개	100%
I-3	없음	없음	-
I-4	중과제 18개	중과제 0개(세부과제 1개)	0%
I-5	세부과제 34개	중과제 2개(세부과제 1개)	(중과제 없음)
I-7	중과제 14개	중과제 14개	100%
I-8	중과제 6개	중과제 6개(세부과제 17개)	100%
I-9	중과제 16개	중과제 16개	100%
I-11	없음	없음	-
I-12	세부과제 61개	세부과제 61개	(중과제 없음)
I-13	없음	없음	-
I-14	중과제 6개	중과제 6개(세부과제 15개)	100%
I-16	중과제 9개(세부과제 107개)	없음	0%
I-17	중과제 6개	중과제 9개(세부과제 23개)	100%
I-19	중과제 9개	중과제 9개(세부과제 89개)	100%
I-20	중과제 15개(세부과제 107개)	중과제 15개(세부과제 83개)	100%
I-21	중과제 1개(세부과제 2개)	중과제 1개(세부과제 2개)	100%
I-22	중과제 13개	중과제 13개	100%
I-23	없음	없음	-
I-24	중과제 11개(세부과제 42개)	중과제 11개(세부과제 42개)	100%
I-25	중과제 3개	중과제 3개	100%

※ *2022년도 설문조사 응답 기준*

해 연구데이터에 대한 관리가 필요한 과제의 수를 함께 조사하였는데, 연구데이터 관리가 필요한 과제란 논문, 특허 등에 필요한 데이터가 R&D를 통해 만들어지는 연구과제를 의미하며, 기술사업화, 실증화 사업 등과 같은 과제는 제외한다. 분석 결과 연구데이터에 대한 관리가 필요한 과제의 수는 중과제 기준으로 총 463개이며, 이는 연구데이터를 생산하는 중과제 488개 중 약 95%를 차지한다. 이를 통해 많은 기관에서 연구데이터의 체계적 관리가 필요한 실정임을 알 수 있다.

3.2 기관별 연구데이터 관리 정책 현황

3.1장에서 출연(연)에서 생산되고 있는 연구데이터의 체계적 관리가 필요한 실정임을 파악한 바 있다. 이 장에서는 각 출연(연)별로 연구데이터 관리 정책이 어떻게 구축되어 있는지 현황을 조사하였다. 다음 〈표 3〉은 각 기관별 DMP 규정 또는 지침의 유무 및 지침 제정 계획 유무

〈표 3〉 기관별 DMP 규정/지침 유무 및 데이터 관리전담인력/조직 유무

기관 ID	DMP 규정/ 지침 유무	DMP 규정/ 지침 제정 계획 유무	데이터 관리전담인력 및 조직의 유무	데이터 관리 주체
I-1	○	-	○	부서
I-3	×	×	×	-
I-4	×	×	○	부서
I-5	×	○	○	개인
I-7	○	-	×	기관
I-8	○	-	×	부서
I-9	○	-	×	기관 / 개인
I-11	×	×	×	개인
I-12	○	-	×	개인
I-13	×	○	×	-
I-14	×	○	○	개인
I-16	○	-	○	기관 / 부서 / 개인
I-17	×	○	×	부서
I-19	×	×	×	개인
I-20	×	×	×	부서
I-21	×	×	○	부서
I-22	○	-	×	부서
I-23	×	×	×	부서
I-24	×	○	×	개인
I-25	○	-	○	기관 / 부서
계	40% (8/20)	42% (5/12)	35% (7/20)	개인 44% (8/18) 부서 56% (10/18) 기관 22% (4/18) ※ *복수응답 포함*

※ *2022년도 설문조사 응답 기준*

와 데이터 관리전담인력과 조직의 유무 및 데이터 관리 주체를 보여준다. DMP는 연구데이터의 생산・보존・관리 및 공동 활용 등에 관한 계획을 의미하며, 국가연구개발과제 제안시 계획서와 함께 별도 문서로 제출하는 것이 일반적인 형태이다. DMP를 작성함으로써 연구수행 전에 데이터에 대한 생산・관리・공유 및 공동 활용에 대한 계획을 수립하게 함으로써 데이터 품질 및 재사용을 제고할 수 있게 할 뿐만 아니라 국가 또는 기관 차원에서 수행된 연구개발과제로부터 생산된 연구데이터의 현황을 파악할 수 있게 하는 핵심적 기능을 수행할 수 있다(최명석, 이상환, 2020).

총 20개의 기관 가운데 절반도 되지 않는 8개의 기관에서만 DMP 규정을 갖고 있었으며, 현재 DMP 지침을 갖고 있지 않는 12개의 기관 가운데 약 42%인 5개의 기관에서 지침 제정 계획을 갖고 있는 것으로 나타났다. 결과적으로 설문에 응답한 총 20개의 출연(연) 기관 가운데 65%인 13개의 기관에서 DMP규정을 갖고 있거나 지침 제정 계획을 갖고 있는 것으로 파악되었다.

DMP의 작성은 앞서 언급한 바와 같이 연구데이터의 현황을 파악함으로써 관리를 용이하게 할 수 있는 역할을 할 수 있다. 연구데이터의 관리가 체계적으로 수행되기 위해서는 추가적으로 전담인력 및 조직이 필수적으로 구축되어야 한다. 데이터 관리전담인력 또는 조직을 갖고 있는 기관은 총 20개 출연(연) 가운데 7개의 기관으로 35%에 불과하고, 대부분의 경우 개인 혹은 부서에서 개별적으로 데이터 관리를 하고 있는 것으로 나타났다. 좀 더 자세히 살펴보면 20개 출연(연) 중 30%인 6개 기관에서는 개인 연구자들이 개별적으로 연구데이터를 관리하고 있는 것으로 나타났고, 40%인 8개 기관에서는 부서 내부적으로 해당 부서에서 생산되는 연구데이터를 관리하고 있는 것으로 파악되었다. I-7의 한 기관에서만 기관차원에서 연구데이터를 관리하고 있다고 응답하였으며, I-9 기관에서는 기관과 개인이, I-25에서는 기관 및 부서에서, I-16은 기관, 부서, 개인 차원에서 연구데이터를 관리하고 있다고 답하였다.

이는 단일 기관 내에서도 연구데이터가 산발적으로 분산되어 개인적 혹은 부서별로 관리되고 있는 것으로 해석할 수 있으며, 일원화된 관리 체계의 부재로 인한 체계적 관리가 어려운 상황임을 유추할 수 있다.

3.3 기관별 연구데이터 관리시스템 현황

연구데이터의 체계적인 관리는 개개인이 수행하기에는 어려운 점이 많이 존재하며, 전담조직이 존재한다 하더라도 전자적 형태의 규모있는 데이터를 효율적으로 관리하기 위해서는 관리시스템이 구축되어야 할 필요가 있다. 이 장에서는 출연(연)의 연구데이터 관리시스템 현황 및 데이터 공개 현황에 대해 분석하였다. 출연(연)별 데이터 관리 전산시스템의 유무는 다음 〈표 4〉와 같다.

총 20개의 기관 가운데 65%인 13개의 기관에서 기관별 데이터 관리 전산시스템을 구축하고 있는 것으로 나타났는데, 데이터 관리 전산시스템을 구축하고 있는 기관 가운데서는 40%도 되지 않는 5개의 기관에서만 해당 전산시스템을 대외 공개중인 것으로 파악되었다. 과반이 넘는 기관에서는 데이터 관리 전산시스템을

〈표 4〉 기관별 데이터 관리 전산시스템 유무 및 공개 범위

기관 ID	데이터 관리 전산시스템 유무	공개 범위
I-1	×	-
I-3	×	-
I-4	○	대외공개
I-5	○	대외공개
I-7	○	기관내부
I-8	○	기관내부
I-9	○	기관내부
I-11	×	-
I-12	○	기관내부
I-13	×	-
I-14	○	기관내부
I-16	○	대외공개 / 기관내부
I-17	×	-
I-19	○	대외공개 / 기관내부
I-20	○	기관내부
I-21	○	대외공개
I-22	○	기관내부
I-23	×	-
I-24	×	-
I-25	○	기관내부
계	65% (13/20)	대외공개 38% (5/13) 기관내부 공개 77% (10/13) ※ *복수응답 포함*

※ *2022년도 설문조사 응답 기준*

기관 내부에서만 공개하여 관리하고 있음을 알 수 있다. 데이터 관리 전산시스템은 데이터 리포지토리와 연계하여 관리할 수 있다는 측면에서도 의미를 갖는데, 데이터 리포지토리는 1차적으로 데이터를 체계적으로 관리 및 보존할 수 있다는 측면에서 가치가 있으며, 2차적으로는 데이터의 공유・활용의 측면에서 유용하게 활용될 수 있다. 리포지토리 이용자들은 해당 리포지토리에 접근함으로써 데이터를 검색 및 접근하여 재활용할 수 있다. 기관 측면에서는 해당 기관의 데이터 관리를 위한 유연한 인프라 구축이 가능하며, 기타 내부 시스템과의 연동을 통해 용이한 데이터 관리가 가능하다. 뿐만 아니라 기관 리포지토리를 DataON과 연계함으로써 기관의 데이터 홍보 및 데이터 접근성 향상을 기대할 수 있다.

각 출연(연)별 데이터 리포지토리의 운영 여부 및 해당 리포지토리와 DataON과의 연계 여부를 정리하면 다음 〈표 5〉와 같다. 설문에 응답한 총 20개의 출연(연) 가운데 연구데이터 리포지토리를 운영하는 기관은 총 11개 기관이며, 그 중 DataON에서 배포가능한 소프트웨어인 연구데이터 저장소(National Research Data Archives, 이하 NaRDA)가 보급된 기관은 총

〈표 5〉 기관별 데이터 리포지토리 운영 여부 및 DataON과의 연계 여부

기관 ID	데이터 리포지토리 운영 여부	NaRDA 배포 여부	데이터 리포지토리 구축 계획 여부	DataON과의 연계 여부
I-1	○	×	-	×
I-3	○	×	-	×
I-4	○	×	-	○
I-5	×	×	×	×
I-7	○	○	-	○
I-8	×	×	○	×
I-9	○	○	-	추진 중
I-11	×	×	보류	×
I-12	○	○	-	○
I-13	×	×	○	×
I-14	○	×	-	×
I-16	○	○	-	○
I-17	×	×	보류	×
I-19	○	×	-	추진 중
I-20	○	○	-	추진 중
I-21	×	○	×	×
I-22	×	×	×	×
I-23	×	×	○	×
I-24	×	×	×	×
I-25	○	○	-	×
계	55% (11/20)	35% (7/20)	43% (3/7)	33% (4/12)

※ 2022년도 설문조사 응답 기준

7개이다. 한편, DataON과의 연계를 위해서는 다음과 같은 부분들이 충족되어야 한다. 먼저 해당 기관은 데이터 리포지토리를 구축 및 운영해야 한다. 추가로 해당 데이터 리포지토리에서 Restful API 또는 OAI-PMH와 같은 프로토콜을 지원함으로써 자동적으로 DataON과 연계 및 수집을 지속적으로 가능케 해야한다. DataON은 데이터의 수집을 위해 DataON의 메타데이터 스키마를 정의하고 있으며,[1] 각 기관에서 구축한 리포지토리의 메타데이터 스키마와 매핑 작업을 통해 연계를 진행하고 있다. DataON에서 정의하고 있는 메타데이터에 포함되지 않는 메타데이터 요소들을 제공받는다면, 해당 요소들은 출연(연) 특성에 맞는 고유 정보로 판단하여, '특성 정보'라는 명칭으로 관리 및 서비스된다. 또한, 출연(연)의 내부 사정으로 인해 대외적으로 공개되는 웹 서비스를 제공하지 못하는 경우에는 DataON에서만 해당 기관의 리포지토리에 접근할 수 있는 권한을 제공받아 메타데이터만 수집하는 체계로 운영 중이다. DataON과의 연계를 활용한 데이터 개방을 진행하기 위해서는 연구데이터 관리 정

1) https://dataon.kisti.re.kr/catalog_summary.do

책 및 조직이 운영되어야 하며, 이와 같은 부분들이 진행되고 있는 출연(연)의 경우에만 아래의 〈표 5〉에 '추진 중'으로 표기하였다. 즉, 연구데이터의 관리를 위해서는 해당 출연(연)이 정책, 인력, 시스템의 세 부분을 모두 구성 및 운영하여야 연구데이터 관리 체계 구축을 위한 준비가 되었다고 판단할 수 있다.

총 20개의 기관 가운데 데이터 리포지토리를 운영하고 있는 출연(연)은 12개 기관이며(60%), 현재 데이터 리포지토리를 운영하고 있지 않은 출연(연) 7개 기관 중에서도 3곳은 데이터 리포지토리 운영 계획을 갖고 있는 것으로 파악되었다. 총 20개 기관 가운데 75%인 15개의 출연(연)이 데이터 리포지토리를 운영 중이거나 구축할 계획을 갖고 있으며, 이미 데이터 리포지토리를 운영하고 있는 12개의 기관 가운데 4개 기관에서는 이미 DataON과 연계를 진행함으로써 연구데이터의 공유・활용 증대를 위한 생태계 구축에 참여하고 있는 것으로 파악되었다.

연구데이터가 적극적으로 활용 및 재활용되기 위해서는 공유 및 공개가 선행되어야 하는데, 총 20개 기관에서 423개의 연구데이터를 보유하고 있는 것으로 조사되었으며, 이 중 개방 중인 데이터는 18건, 개방 가능 데이터는 83건, 내부 개방 데이터 1건, 관련자에게만 개방 데이터 1건, 미개방 데이터가 295건인 것으로 조사되었다. 각 기관별 개방 및 개방 가능 데이터 건수와 비율을 정리하면 다음 〈표 6〉과 같다.

〈표 6〉 기관별 보유 연구데이터 수 및 개방 가능 데이터 수

기관 ID	총 보유 연구데이터 수	개방 여부	개방 가능 데이터 비율
I-1	1건	-	0%
I-3	15건	-	0%
I-4	3건	개방 중 3건	100%
I-5	1건	개방 가능 1건	100%
I-7	62건	개방 가능 62건	100%
I-8	2건	내부 개방 1건, 미개방 1건	0%
I-9	24건	개방 가능 20건, 미개방 4건	83%
I-11	0건	-	0%
I-12	261건	미개방 261건	0%
I-13	1건	개방 중 1건	100%
I-14	1건	-	0%
I-16	9건	미개방 9건	0%
I-17	23건	개방 중 14건, 미개방 9건	60%
I-19	10건	미개방 10건	0%
I-20	0건	-	0%
I-21	0건	-	0%
I-22	2건	(미응답)	0%
I-23	0건	-	0%
I-24	6건	-	0%
I-25	2건	관련자 개방 1건, 미개방 1건	50%

※ 2023년도 설문조사 응답 기준(2023년도 미응답 기관의 경우에만 2022년도 기준)

설문에 응답한 총 20개의 기관 가운데 65%인 13개의 기관에서는 공개할 수 있는 데이터가 없다고 응답하였으며, 4개의 기관에서는 모든 데이터가 개방 중이거나 개방 가능한 데이터라고 응답하였고, 3개의 기관에서는 절반 이상의 보유 데이터를 개방 중이거나 개방 가능하다고 답하였다. 과반이 넘는 출연(연)에서 기관의 보유 데이터를 공개하기 어렵다고 응답하였는데 비공개 사유로는 보안 문제, 보상 체계의 부재, 기업 데이터의 혼재, 공개 시점의 미도래, 국가 경쟁력에 영향을 미칠 수 있는 데이터 등으로 파악되었다.

3.4 연구데이터 관리 현장의 어려움

본 연구에서는 기관별 연구데이터 관리 현황과 더불어 연구데이터 관리 측면에서 어려운 점에 대한 질문을 실시하였다. 설문에 응답한 연구데이터 관리자들의 응답은 기관별로 다양하게 나타나는 경향이 있지만, 크게 네 가지의 측면으로 정리하면 다음과 같다. 첫째, 연구데이터 관리와 관련한 인력 및 예산의 부족이다. 현장에서 가장 어렵게 느껴지는 부분은 연구데이터의 수집, 관리 및 활용을 담당하여 수행할 수 있는 관리 인력 부족의 문제임이 파악되었다. 많은 기관에서 연구데이터 관리 부분에 대한 투자가 적은 상황을 지적하면서 연구데이터의 수집, 관리 및 활용을 위한 리포지토리 또는 플랫폼의 개발 과정에서 충분한 예산 및 인력 확보에 어려움이 존재하며, 이와 관련한 정부의 지원의 필요하다고 응답하였다.

둘째, 연구데이터 관리에 대한 정책의 부재가 현장에서 겪는 어려움으로 파악되었다. 연구데이터 관리를 위한 인력 및 예산 확충을 위한 정부의 지원이 필요함과 동시에 이를 위한 법적 근거가 확충될 필요가 있다는 의견이 다수 개진되었다. 이는 정부 차원에서 연구데이터의 공유 및 활성화를 위한 정책이 수립되는 것이 바람직할 것이며, 특히나 정부 과제의 경우에는 연구데이터 및 연구 성과물을 종합적으로 검색할 수 있는 플랫폼이 구축되어 어떤 분야의 연구자라도 해당 데이터를 손쉽게 요청할 수 있도록 하는 것이 활용 측면에서 바람직할 것이라는 의견도 나타났다.

셋째, 연구데이터 공개 시 연구자에게 주어지는 보상체계의 부재 역시 연구데이터 공개를 저해하는 요인으로 파악되었다. 현장에서는 현재 연구데이터의 등록을 개별 연구자에게만 일임하고 있는 현상에 대한 문제점을 지적하였는데, 연구데이터의 등록 및 관리를 단순히 정량적・정성적으로만 평가하기에는 제도 구축과 공감대 형성이 부족한 실정이며, 연구데이터를 등록 시 필요한 연구자들의 노력은 특히 논문, 특허, 기술이전 등과 같이 일반적으로 연구자들이 받는 평가와는 거리가 멀기 때문에 연구데이터를 등록 및 공개하였을 때 연구자들에게 주어지는 보상책 마련 근거가 필요하다는 의견이 다수 존재하였다.

마지막으로, 연구데이터 특성에 대한 고려가 부족한 상황에 대한 문제점이 지적되었다. 연구데이터는 데이터 자체의 특성상 개인정보 및 생명윤리 관련 등에 대한 내용을 포함하고 있는 경우가 있어서 적극적으로 공개하기 어려운 데이터가 다수 존재하기도 한다. 그렇기 때문에 이와 같은 특성을 고려하여 일괄적인 공개 및 공유 요구가 아닌 각 연구데이터의 특성에 맞는 공개 기

준이 필요할 것이라는 의견이 파악되었다.

4. 연구데이터 공유 체계 구축 방안

4장에서는 앞서 진행한 출연(연) 보유 연구데이터 현황, 연구데이터 관리 정책 현황, 연구데이터 관리시스템 현황 및 연구데이터 관리 현장의 어려움에 관한 분석을 기반으로, 연구데이터 공유 및 관리 체계 구축을 위한 방법을 정책적 측면, 시스템적 측면, 연계 체계 구축 측면으로 나누어 제언한다.

4.1 일원화된 관리 정책 구축

출연(연) 별로 연구 분야의 특성에 따른 연구데이터 정의 및 범위 설정, 관리 지침, 관리 시스템, 구축 비용, 전담 인력 등의 복합적인 요소로 인하여 관리 체계가 상이하다. 연구데이터의 수집 및 관리에서 연구데이터 활용으로 중요성이 증대되는 시점에서 각 출연(연)마다 연구데이터를 수집, 저장, 관리, 공개와 공유하는 관리 정책을 별도로 수립하는 것보다는 일원화된 관리 정책을 구축하는 것이 필요할 것이다. 이는 과학기술분야의 연구데이터를 체계적으로 활용하는 여건을 조성하고, 연구성과의 축적과 공유를 통해 데이터 기반의 연구개발 환경을 구축하는 과학기술정보통신부의 국정과제 '국가혁신을 위한 과학기술 시스템 재설계' 추진에도 부합하는 방향이다.

현재 국내・외 국가기관 및 연구데이터플랫폼 별로 연구데이터 전반에 관한 관리 지침을 제공하고 있으나, 국제 표준과 같이 공통으로 참조할 수 있는 모델은 아직 미흡한 실정이다. 일반적으로 연구데이터를 관리하고, 활용하기 위한 공정한 접근 및 이용을 보장하는 FAIR 정책을 기반으로 관리 지침을 수립하고 있다. FAIR 정책은 검색가능성(Findable), 접근가능성(Accessible), 상호운용성(Interoperable), 재사용성(Reusable)을 보장해야 하는 정책으로 연구데이터의 보호와 활용이 공존하기 위한 규칙을 포함한다.

이와 함께, 연구데이터 관리 시스템을 대상으로 연구데이터의 수집, 보존, 관리, 개방, 공유 등을 포함한 항목을 평가하여 인증하는 제도도 존재하는데, 연구데이터의 관리 체계가 적절한지 판단하는 것으로 CoreTrustSeal(이하 CTS)이 대표적이다. CTS는 2017년에 국제연구데이터연맹(Research Data Alliance, 이하 RDA) 산하의 세계데이터시스템(World Data System, 이하 WDS)과 Data Seal of Approval (DSA)이 출범한 인증기관으로, 인프라의 지속 가능성을 위한 신뢰할 수 있는 연구데이터 관리 시스템을 인증한다. 연구데이터 관리 시스템을 17개의 요구사항으로 상세하게 구분하여 종합적인 판단하에 인증하고 있으며, 24년 3월에 국내에서는 DataON이 최초로 인증을 획득한 바 있다.

일원화된 관리 정책이 출연(연)의 연구현장에 반영되기 위해서는 연구데이터 관리 지침과 함께 법제도로 규정하는 것이 더 효과적이다. 아직까지 국가연구개발사업에서 발생하는 연구데이터를 관리하기 위해 해당 내용을 법으로 규정하는 국가는 없는 실정이다. 미국의 경우, 22년에 대통령 메모랜덤을 통해 연방정부의 지원을 받은 과제의 출판물과 연구데이터의 접근성을 제고하고 책임 있는 관리 및 사용이 시행

되어야 한다는 것을 명시한 바 있다. 유럽연합은 21년도 Horizon Europe을 통해 연구데이터의 접근과 상호운용, 재사용을 보장하도록 하고 있으며, 같은 해에 프랑스는 제2차 오픈사이언스 계획을 행정계획으로 제시하고, 호주는 책임 연구 수행 강령과 세부 지침으로, 일본은 제6차 과학기술혁신기본계획에 공적지원한 연구데이터의 관리・활용 기본방향을 포함하였다. 독일은 추후 연구를 위한 데이터 접근 법률로 연구데이터법을 추진할 예정이다.

우리나라의 경우, 2023년 9월에 과학기술정보통신부에서 「국가연구데이터 관리 및 활용 촉진에 관한 법률안」(이하 연구데이터법)을 입법예고하였고, 당해 11월에 공청회를 실시하여 연구현장의 혼란을 최소화하고, 애로사항을 반영하려는 노력을 기울인 바 있다. 연구데이터법은 연구데이터를 체계적으로 관리하기 위한 다양한 시도는 있으나 법제도 인프라의 미비로 인하여 활성화에 한계를 인식하고 국가연구개발 과정에서 생산되는 연구데이터를 체계적으로 관리하기 위한 근거법이다. 연구데이터법은 국가연구데이터의 정의, 적용범위, 수집・관리, 보유, 등록・공개, 활용, 통합플랫폼, 선도기관, 분야별 전문센터, 분쟁조정 등의 내용을 포함하고 있어 출연(연)이 일원화된 관리 정책을 기반으로 연구데이터를 체계적으로 관리할 수 있도록 법제도적인 측면에서 근거로 활용될 수 있을 것으로 기대한다. 이는 현장에서 연구데이터 관리 시 애로사항으로 지적되었던 정부차원의 지원 문제를 해결할 수 있는 부분이 될 것이다. 보다 구체적으로, 연구데이터의 법제도 정책 제안은 연구데이터법, 시행령, 기관별 연구데이터 관리지침 등으로 구분되어 있으며, 연구현장의 의견을 반영하기 위한 공청회를 통해 연구데이터의 관리, 활용, 보안, 주권, 인센티브 방안을 제안할 수 있다.

4.2 데이터 관리 및 공유를 위한 기관 리포지토리 구축

연구데이터의 효율적인 관리, 보존 및 활용을 위해서는 리포지토리 및 관리시스템 구축이 필요하다. 그러나 본 연구에서 수행한 설문조사에 따르면 1/3에 가까운 수의 출연(연)에서 데이터 관리 전산시스템 없이 부서 또는 개인이 연구데이터를 관리하고 있는 것으로 나타났으며, 비슷한 수의 출연(연)이 리포지토리를 구축하고 있지 않은 것으로 파악되었다. 데이터 리포지토리 및 관리시스템은 단순히 기관의 데이터를 통합관리 및 보존한다는 측면 외에도 데이터 공유, 활용 및 재활용의 측면에서도 가치를 갖는다. 수집, 보존, 공개 및 활용되는 연구데이터는 전자적 형태를 띄고 있으며, 전자적 형태의 파일은 손상 및 손실의 위험 등에 대비하기 위해 데이터 형태에 맞는 관리시스템 및 리포지토리를 통해 관리 및 보존되는 것이 요구된다.

뿐만 아니라 데이터 리포지토리 및 관리시스템을 활용하여 각 연구데이터 특성에 맞는 관리가 가능하다. 본 연구에서는 데이터 관리자들로부터 개인정보를 포함하고 있는 특정 연구데이터와 같이 개방 및 관리에 대한 기준을 일괄 적용하기 어려운 경우를 파악하였다. 기관별 리포지토리에 데이터를 등록 시 데이터 접근 및 활용에 대한 권한 설정, 보존 수준, 라이센스 등에 대한 내용을 명시하고 수준별로 적용함으로써

체계적이고 적극적으로 연구데이터를 관리 및 활용할 수 있음을 기대할 수 있다.

연구데이터는 연구 과정에서 생산되는 데이터이기 때문에 여타 다른 데이터와 비교할 때 연구 내용의 전문성을 특징으로 가질 수 밖에 없는 특성을 갖는다. 해당 데이터의 전문성을 유지하는 수준의 품질 관리를 진행하기 위해서는 데이터를 생산하는 기관의 측면에서 직접 데이터 리포지토리를 운영하고 관리하는 것이 바람직할 수 있다. 그러나 이용자의 입장에서는 연구데이터를 검색 및 접근하기 위한 다양한 리포지토리가 존재하는 것은 비효율적인 접근 방법일 수 있다. 그렇기 때문에 각 연구 분야 및 기관의 특성에 맞는 리포지토리를 구축하되 통합 접근성 및 일원화된 서비스를 제공할 수 있도록 통합 플랫폼에 연계하는 방향이 바람직할 것이다. 국내에는 DataON이 국내 출연(연)을 대상으로 국가연구개발과제에서 생산되는 연구데이터를 통합적으로 수집하여 제공할 수 있도록 노력하고 있으며, 이와 같은 방법에 대한 인식 증대는 이용자들로 하여금 통합 플랫폼에 대한 접근성을 향상시킬 수 있을 것이며 향상된 접근성은 보다 활발한 연구데이터의 활용 및 재활용으로 이어질 수 있을 것이다.

연구데이터의 효율적인 관리 및 공유를 위한 기관 리포지토리를 구축하고 지속적으로 운영하기 위해서는 다음과 같은 내용을 만족시켜야 할 필요가 있다. 첫째, 국가 정책의 방향에 맞춘 기관별 연구데이터 관리 정책을 확립하여야 한다. 각 기관에서 생성 및 관리하고 있는 연구데이터의 정의를 기본으로 하여 데이터 수집 및 보존 정책, 공개 범위 설정, 데이터 권한 설정, 보안 수준, 라이센스 등에 대한 기관별 관리 정책을 명시하고 이를 기준으로 데이터 리포지토리를 운영하여야 한다. 둘째, 연구데이터를 전담으로 관리하는 부서 및 인력을 배치할 필요가 있다. 본 연구에서 분석한 내용에 따르면 대다수의 출연(연)이 연구데이터를 관리하는 전담부서 및 인력의 부재로 어려움을 겪고 있는 것으로 파악되었다. 연구데이터의 효율적 관리를 위한 기관 리포지토리의 구축 및 지속적이고 안정적인 운영을 위해서는 연구데이터에 대한 이해도가 높은 인력을 포함하는 전담부서를 배치하여야 한다. 셋째, 시스템적인 측면에서 기술적(technical) 인프라를 구축해야 한다. 기관 리포지토리는 전자적 형태를 띄는 데이터를 수집, 관리, 보존 및 배포하는 온라인 저장소이다. 이는 물리적 저장소와는 다른 성격을 보일 수 밖에 없으며 이를 성공적으로 운영하기 위해서는 연구데이터에 대한 이해 뿐만 아니라 리포지토리 자체에 대한 기술적인 이해가 바탕이 되어야 한다. 재정 및 인력의 부족으로 기관 내부에서 직접 리포지토리를 구축하기 어려운 경우에는 NaRDA와 같은 소프트웨어를 활용하여 이를 기관 특성에 맞게 커스터마이즈하여 이용할 수도 있다. 넷째로, 공개 기준 및 공유 범위를 확정하여 기관 리포지토리의 성격을 정의할 필요가 있다. 기관 내부로 활용할 것인지, 외부로 공개할 것인지 등에 대한 공개 범위를 설정한 후에 DataON과 같은 통합플랫폼과의 연계 수준을 정의하여야 한다. 연구데이터가 공개 및 공유된 후에는 연구데이터의 재사용과 같은 활용 내역을 추적하여 해당 연구데이터를 공개한 연구자에게 제공할 수 있는 보상 내역 등을 고민하는 것이 바람직할 것이다.

4.3 DataON을 중심으로 한 연계 체제 구축

DataON은 2018년 국가과학기술심의위원회에서 의결된 연구데이터 공유・활용 전략(안)에서 연구데이터를 원스톱으로 검색에서부터 활용에 이르기까지의 환경을 제공하는 플랫폼으로 가장 처음 정의하였다. 전략(안)에서는 인공지능, 빅데이터 기술의 발전에서 가장 중요한 것은 데이터임을 강조하였으며, 전략(안)에서 언급한 바와 같이 현재 GPT와 같은 다양한 모델의 등장으로 급속하게 발전하고 있는 인공지능 기술은 모델을 생성하기 위해 필요한 데이터가 충분히 존재했기 때문에 실현 가능한 방법론이라고 할 수 있다. 이처럼 인공지능의 기술은 어떤 데이터를 어느 정도로 수집하여 활용해야 하는지 정확하게 파악할 수 없는 상황이기 때문에, 국가 연구 자산의 한 종류로써 데이터 수집부터 활용까지 보다 체계적으로 관리할 필요성이 존재한다.

이에 따라 현재까지 DataON에서는 현재 총 7개의 출연(연)의 연구데이터 시스템과 연계하고 있다. 출연(연) 연계의 경우, 메타데이터를 중심으로 연계하고 있으며, 원본 데이터는 각 출연(연) 연구데이터 시스템의 랜딩 페이지 또는 별도 연락을 통한 데이터 활용 절차를 두고 운영되고 있다. DataON에서는 OpenAPI 호출과 웹 키워드 검색을 통해 연구데이터를 조회할 수 있는 기능을 제공하고 있으며, 이를 통해 조회한 데이터의 건수는 월 평균 약 36만 8천건 정도이다. 또한, DataON에서는 우리나라 출연(연)에서 산출되는 연구데이터를 한 곳에 모아 검색할 수 있도록 지원함으로써 정책입안자의 경우에는 우리나라 R&D 예산을 투자하여 생산되는 산출물 중 하나인 연구데이터를 관리할 수 있는 토대를 마련할 수 있고, 연구자들은 필요한 데이터를 어디에서 얻을 수 있는지를 한 사이트에서 찾을 수 있음으로써 데이터 탐색에 필요한 시간과 비용을 줄일 수 있다는 장점을 갖는다. DataON은 2026년도까지 전체 출연(연)의 연구데이터를 연계하는 것을 목표로 하고 있으나 아직 연구데이터 관리 체계가 마련되지 않은 출연(연)이 상당 수 존재하기 때문에 해당 기관들을 위해, 데이터 관리 및 공유를 위한 기관 리포지토리인 NaRDA를 개발 및 보급함으로써 데이터 관리 및 공유 체계를 출연(연)에서 먼저 구축할 수 있도록 지원하고 있다. 본 연구의 설문조사 분석 결과에서도 나타난 바와 같이 기관에서 연구데이터 관리 체계가 구축되기 위해서는 데이터 관리 정책 및 전담 인력이 마련되어야 한다. 이를 기관에서 추진하기 위해서는 관련 정책적 근거가 필요하며, 이를 위해 2023년 과기정통부에서는 연구데이터법을 입법예고한 바 있다.

연구데이터법의 경우 아직 법률안이 통과되지 않은 상황이며, 현재는 법제화와 같은 정책적 근거가 부족한 실정이다. 그렇기 때문에 DataON은 국가연구개발정보처리기준에 정의된 수준 및 2018년도 연구데이터 공유・활용 전략(안)에서 제시한 바에 따라 운영되고 있으며, 개별적으로 분야별 데이터 전문 센터가 연구데이터를 수집하고 있다. 아직은 DataON에서 일원화되어 연구데이터를 수집해야 할 근거가 부족하기 때문에 현재는 연계를 원하는 출연(연) 간의 자발적인 노력에 의하여 연계를 확장해 나가고 있다. 따라서 앞으로 정책적 근거가 마련되어 DataON이 국가연구데이터플랫폼으로써

국가 R&D에서 산출되는 연구데이터를 종합적으로 수집할 수 있다면, 연구자에게 국가연구개발분야별로 연구데이터를 검색하여 제공할 수 있는 환경을 제공할 수 있을 뿐만 아니라, 연구데이터를 필요로 하는 수요자에게 관련 분야의 연구기관과의 매칭 등을 통해 융합연구 또는 공동연구에 필요한 데이터 공유 및 활용을 지원할 수 있을 것으로 기대한다.

아울러, DataON에서는 연구데이터를 직접 관리하기 어려운 개인 연구자, 소규모 연구 그룹 또는 연구 기간이 종료된 연구과제에서 생산된 연구데이터를 관리하는 역할 또한 포함하고 있기 때문에, 앞으로 과제 종료 또는 데이터 관리가 어려운 환경으로 인해 소실될 수 있는 국가 R&D 자산을 지속적으로 관리할 수 있는 토대를 마련할 수 있는 기점이 될 수 있다. 또한, 최근 우리나라의 주요한 사회 문제 중 하나로 대두되고 있는 청년 인구의 감소로 인해 많은 대학원 연구실의 운영이 중단되거나 사라지는 추세이다. 이는 기존 연구실 선・후배 사이의 교육을 통해 전수되던 연구 노하우 및 데이터가 더 이상 전달되기 어려운 상황을 의미하는 것으로도 해석할 수 있다. 따라서 연구를 통해 생산된 데이터를 대학교 등의 연구기관 또는 DataON에서 활용할 수 있도록 지원할 수 있다는 것은 연구데이터의 보존 및 관리의 관점에서 큰 의미를 가질 수 있다. 각 출연(연)과 같은 기관에서의 연구데이터 공유 및 관리를 위한 체계 구축도 중요하지만, 각 기관에서 해결하기 어려운 부분들을 지원하고 해당 기관에서 공유하고자 하는 연구데이터를 한 곳에서 모아 검색 및 활용할 수 있도록 해주는 역할을 DataON을 통해 수행해 나갈 수 있을 것으로 기대한다.

5. 결론 및 제언

본 연구는 연구데이터의 공유 및 활용 촉진을 목적으로 NST 산하 출연(연)의 연구데이터 관리 현황을 분석하고, 이를 바탕으로 연구데이터 공유 활용 및 관리 체계 구축을 위한 방법을 제언하였다. 연구데이터 관리 현황 분석은 총 20개의 출연(연)을 대상으로 진행되었으며, 기관별 보유 연구데이터 현황, 기관별 연구데이터 관리 정책 현황, 기관별 연구데이터 관리시스템 현황을 분석하였다. 출연(연) 별로 기관의 특성에 따라 보유 연구데이터 수에는 상당한 차이가 존재하는 것으로 나타났으며, 이는 보유 연구데이터가 없다는 응답에서부터 총 260건이 넘는 연구데이터를 보유하고 있다는 응답까지 매우 다양하게 나타났다. 출연(연)에서는 다양한 국가연구개발과제를 진행하는데, 각 기관별로 다양한 사업을 통해 연구데이터를 생산해내고 있었으며, 이 중 약 95%의 사업에 대해 연구데이터에 대한 관리를 필수적으로 진행해야 함을 알 수 있었다. 이렇듯 생산해내는 데이터 중 대부분의 데이터에 대한 체계적 관리가 필요함에도 불구하고 현재 조사 대상인 출연(연) 가운데 절반도 되지 않은 기관에서만 DMP 규정을 갖고 있는 것으로 확인되었으며, 기관별로 데이터 관리전담인력 또는 조직을 갖고 있는 기관 역시 겨우 35%에 불과한 실정으로 나타났다. 현재는 많은 기관에서 일원화된 관리 체계의 부재로 인해 연구데이터의 체계적인 관리가 어려운 상황이라고 볼 수 있다. 뿐만 아니라 다양한 이유로 연구데이터의 체계적 관리 및 공유에 어려움이 있음을 파악하였는데, 이는 재정적 지원의 부족, 법적・정

책적 근거의 부족, 연구자들을 대상으로 한 보상 체계의 부재 등을 포함한다.

본 연구에서는 위와 같은 현황 분석을 기반으로 연구데이터 공유 체계 구축을 위한 다양한 접근점을 제언하였다. 먼저 정책적인 근거를 마련하여 연구데이터의 정의, 적용 범위, 관리의 내용, 활용 방식, 선도기관 등의 내용을 제시하는 것이다. 연구데이터법과 같은 정책적 제도가 뒷받침되면 전담 조직의 구성 및 재정적 지원과 같은 문제점에 대한 해결 방안으로 작용할 수 있을 것일 뿐만 아니라 연구자들의 인식 전환에도 도움이 될 것으로 기대한다. 또한 기관별 리포지토리 및 관리시스템 구축을 통해 전자적 형태의 데이터를 그 특성에 맞게 체계적으로 관리함과 동시에 DataON과 연계함으로써 이용자에게는 통합 접근성을 제공하는 것이 바람직할 것이다. 연구데이터의 특성상 전문성을 담보하기 위해서는 해당 데이터를 생산해낸 기관에서 품질 등에 대한 관리를 진행할 필요가 존재한다. 그러나 이용자의 입장에서는 산재되어있는 데이터를 각 기관의 리포지토리에 접근하여 검색하는 것은 비효율적일 수 밖에 없다. 그렇기 때문에 기관별로 리포지토리 및 관리시스템을 구축하여 관리하되 이를 DataON과 연계하고 원스톱 서비스를 제공함으로써 데이터의 접근성 및 활용도를 제고할 수 있을 것이다.

추가로, 연구데이터의 정책적 지원 및 실무적 수행 방안을 논의하기 위해 협의체를 구성하여 지속적으로 논의를 진행하는 것이 바람직할 것이다. KISTI는 연구데이터법의 제정과 출연(연) 적용에 관한 연구현장 의견 수렴을 위한 출연(연) 협의체를 이미 23년도부터 운영하고 있으며, 이를 통해 연구데이터 개방 및 활성화를 촉진할 수 있음을 기대할 수 있다. 뿐만 아니라 이와 같은 정책적 근거 마련과 현장 적용에 대한 의견 수렴을 위한 협의체 외에도 출연(연) 연구데이터 관리자로 구성된 실무협의체를 구축함으로써 실무부서에서 개방가능한 연구데이터를 지속적으로 발굴하고 현장의 의견을 들을 수 있을 것이다. 출연(연)에서 생산 및 관리되는 연구데이터의 개방 및 활용 활성화를 위해 각 기관별 연구데이터 실무 담당자로 구성된 실무협의체가 구축되어 연구현장의 이해를 돕고, 데이터 관리의 실제를 파악할 수 있을 것으로 기대한다.

본 연구는 국내 출연(연) 연구데이터 관리 현황을 분석함으로써 현재 생산되고 있는 연구데이터가 어떻게 관리 및 공유되고 있는지, 관리 및 공유에 있어서 어려움과 개선되어야 할 점은 무엇인지 파악하고, 이를 기반으로 연구데이터 공유 및 관리 체계 구축을 위한 방법을 정책적 측면, 시스템적 측면, 연계 체제 구축 측면으로 나누어 제언한다는데 의의를 갖는다. 그러나 조사 분석 대상인 출연(연)이 특정 분야의 연구데이터를 중점적으로 생산한다는 특성상 해당 기관에서 공개하는 데이터의 성격 및 공개 가능한 이유 등에 대한 내용 분석이 제공될 경우 해당 연구데이터를 생산하는 출연(연)에 대한 유추가 가능할 위험이 있기 때문에 각 기관이 관리하고 있는 데이터에 대한 상세한 내용을 담지 못한다는 것이 제한점으로 남는다. 후속 연구를 통하여 각 출연(연)이 관리 및 공유하는 데이터에 대한 보다 상세한 분석을 기반으로 연구데이터의 공유 및 활용 체계 구축에 관한 연구가 진행되기를 기대한다.

참 고 문 헌

과학기술기본법. 법률 제18727호.

과학기술정보통신부 (2020). 국가연구개발정보처리기준 과학기술정보통신부고시 제2020-102호.

국가과학기술연구회 (2019). 연구데이터 관리 가이드라인(2019-07).

김예현, 김지현 (2023). 과학기술분야 기관 연구데이터 리포지터리 운영 활성화 방안 연구. 한국비블리아학회지, 34(3), 109-134.

김주섭, 김선태 (2023). 대학도서관 연구데이터 관리 서비스 현황 및 제안: 과학기술특성화 대학을 중심으로. 한국문헌정보학회지, 57(3), 279-301. https://doi.org/10.4275/KSLIS.2023.57.3.279

김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455. https://doi.org/10.16981/kliss.43.3.201209.433

김지현 (2015). 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 한국문헌정보학회지, 49(3), 413-436. https://doi.org/10.4275/KSLIS.2015.49.3.413

박성은, 이미경, 조민희, 송사광, 김다솔, 임형준 (2024). 국가연구데이터커먼즈 체계 수립을 위한 연구데이터 관리자들의 인식에 관한 연구. 정보관리학회지, 41(1), 465-486. https://doi.org/10.3743/KOSIM.2024.41.1.465

최명석, 이상환 (2020). 데이터 관리 계획의 국내 현황 및 과제. 한국콘텐츠학회 논문지, 20(6), 220-229. https://doi.org/10.5392/JKCA.2020.20.06.220

한국과학기술정보연구원 (2019). 연구데이터 공유확산체제 구축(K-19-L01-C03).

한나은 (2022a). 연구데이터 공유 및 활용을 위한 전문가 실천공동체 구축에 관한 연구. 한국문헌정보학회지, 56(4), 181-203. https://doi.org/10.4275/KSLIS.2022.56.4.181

한나은 (2022b). 연구데이터 큐레이터로 구성된 실천공동체 구축에 관한 질적 연구. 한국문헌정보학회지, 56(3), 119-152. https://doi.org/10.4275/KSLIS.2022.56.3.119

한나은, 김성희 (2014). 외국 대학도서관의 디지털 큐레이션 프로세스 비교분석. 한국도서관·정보학회지, 45(2), 93-116. https://doi.org/10.16981/kliss.45.2.201406.93

황미녕, 박상배, 신영호, 신성호, 서동민 (2023). 연구데이터 공유·관리·분석을 위한 국가연구데이터 플랫폼(DataON) 서비스 운영. 한국콘텐츠학회 논문지, 23(1), 36-47. https://doi.org/10.5392/JKCA.2023.23.01.036

National Science Foundation (2014). Proposal and Award Policies and Procedures Guide (nsf 15001).

Pilat, D. & Fukasaku, Y. (2007). OECD principles and guidelines for access to research data from public funding. Data Science Journal, 6, OD4-OD11. https://doi.org/10.2481/dsj.6.OD4

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide. PloS one, 15(3), e0229003. https://doi.org/10.1371/journal.pone.0229003

Tolle, K. M., Tansley, D. S. W., & Hey, A. J. (2011). The fourth paradigm: data-intensive scientific discovery [point of view]. Proceedings of the IEEE, 99(8), 1334-1337. https://doi.org/10.1109/JPROC.2011.2155130

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Choi, Myung-Seok & Lee, Sang-Hwan (2020). Currents status and issues of data management plan in Korea. The Journal of the Korea Contents Association, 20(6), 220-229. https://doi.org/10.5392/JKCA.2020.20.06.220

Han, Na-Eun & Kim, Seong-Hee (2014). Comparative analysis on digital curation process in foreign academic libraries. Journal of Korean Library and Information Science Society, 45(2), 93-116. https://doi.org/10.16981/kliss.45.2.201406.93

Han, Na-Eun (2022a). A study about building a community of practice of experts for sharing and using research data. Journal of the Korean Society for Library and Information Science, 56(4), 181-203. https://doi.org/10.4275/KSLIS.2022.56.4.181

Han, Na-Eun (2022b). Building a community of practice of research data curations: a qualitative study. Journal of the Korean Society for Library and Information Science, 56(3), 119-152. https://doi.org/10.4275/KSLIS.2022.56.3.119

Hwang, Mi-Nyeong, Park, Sang-Bae, Shin, Young-Ho, Shin, Sung-Ho, & Seo, Dong-Min (2023). DataON service operation for research data sharing, management and analysis. The Journal of the Korea Contents Association, 23(1), 36-47. https://doi.org/10.5392/JKCA.2023.23.01.036

Kim, Ji-Hyun (2012). A study on university researchers' data management practices. The Korea Journal of Library and Information Science, 43(3), 433-455. https://doi.org/10.16981/kliss.43.3.201209.433

Kim, Ji-Hyun (2015). A study on the perceptions of university researchers on data management and sharing. Journal of the Korean Society for Library and Information Science, 49(3), 413-436. https://doi.org/10.4275/KSLIS.2015.49.3.413

Kim, Ju-Seop & Kim, Sun-Tae (2023). Current status and proposal of university library research data management service: focused on science and technology specialized universities. Journal of the Korean Society for Library and Information Science, 57(3), 279-301. https://doi.org/10.4275/KSLIS.2023.57.3.279

Kim, Ye-Hyeon & Kim, Ji-Hyun (2023). A study on strategies to promote the activation of institutional research data repositories in the field of science and technology. Journal of the Korean Biblia Society for Library and Information Science, 34(3), 109-134.

Korea Institute of Science and Technology Information (2019). Establishment of Research Data Sharing and Dissemination System (K-19-L01-C03).

Ministry of Science and ICT (2020). National Research and Development Information Processing Standards Ministry of Science and ICT Notice No. 2020-102.

National Research and Development Innovation Act. Section No18645.

National Research Council of Science and Technology (2019). Research Data Management Guidelines (2019-07).

Park, Seong-Eun, Lee, Mi-Kyoung, Cho, Min-Hee, Song, Sa-Kwang, Kim, Da-Sol, & Yim, Hyung-Jun (2024). A study on the perception of research data managers to establish a Korea research data commons system. Journal of the Korean Society for Information Management, 41(1), 465-486. https://doi.org/10.3743/KOSIM.2024.41.1.465

[부록 1] 설문 문항

Part 1. 기관별 연구데이터 관리 일반 현황

1. 귀 기관에서 운영하고 있는 2022년도 기본사업 중 연구데이터가 생산되는 사업의 개수와 예산 규모를 말씀해 주시기 바랍니다.
   ※ *기관별 기본사업 관리 단위에 따라 세부과제까지 관리하는 기관에서는 중과제, 세부과제로 작성 부탁드리며, 중과제로 관리하는 기관은 중과제에 대한 내용으로 작성 부탁드립니다.*
   답변: 중과제 개 (억원) (세부과제 개)

2. 귀 기관에서는 국가연구개발정보처리기준에 따라 연구데이터 관리가 필요한 과제가 있습니까? 만일 존재하지 않는다면 그 이유를 말씀해 주시기 바랍니다. (해당되는 부분에 ■로 표시)
   ※ *연구데이터 관리가 필요한 과제란?*
   - *논문, 특허 등에 필요한 데이터가 R&D를 통해 만들어지는 연구과제*
   - *기술사업화, 실증화 사업 등과 같은 과제는 제외*

□ 유	□ 무 (사유:)

Part 2. 기관별 연구데이터 관리 정책 현황

3. 귀 기관에서는 연구데이터관리계획에 대한 규정/지침이 있습니까? (해당되는 부분에 ■로 표시)

□ 유	□ 무

4. 귀 기관에서는 연구데이터관리계획 규정/지침을 제정할 계획을 가지고 있습니까? (해당되는 부분에 ■로 표시)

□ 유	□ 무

5. 현재 귀 기관에서 수행하고 있는 주요사업 중에서 국가연구개발정보처리기준에 따라 연구데이터 관리가 필요한 과제의 수는 몇 개가 있습니까?
   답변: 중과제 개 (세부과제 개)

6. 귀 기관에서는 연구데이터를 주로 어디에서 관리하고 있습니까?

□ 기관	□ 부서	□ 개인

7. 귀 기관에서는 연구데이터 또는 데이터 관리에 대한 전담인력이나 조직을 보유하고 있습니까?

□ 유	□ 무

Part 3. 기관별 연구데이터 관리 인프라 현황

8. 연구데이터 관리를 위해 기관 또는 부서에서 운영하는 전산시스템이 있습니까? (해당되는 부분에 ■로 표시)

□ 유 *(8-1번으로 이동)*	□ 무 *(9번으로 이동)*

8-1. 전산시스템이 있다면, 해당 시스템은 대외 공개되어 있습니까? 아니면 기관 내부에서만 운영되고 있습니까? (해당되는 부분에 ■로 표시)

□ 대외공개	□ 기관내부

9. 귀 기관에서 보유하고 있는 주요 데이터에 대하여 기술하여 주시기 바랍니다.
※ 데이터 유형(① 텍스트 ② 동영상 ③ 이미지 ④ 소리 ⑤ 기타)

번호	데이터 명	데이터 개요 (설명)	데이터 유형	크기	특성 (생산주기, 연관 장비 등)	해당 국가과학기술표준분류 (최대 3개)
1						
2						
…						

※ 국가과학기술표준분류체계 개정 고시안(2018) 참조

10. 귀 기관에서 외부 공개 가능한 데이터가 있습니까?

□ 유 *(10-1번으로 이동)*	□ 무 *(11번으로 이동)*

10-1. 공개 가능한 데이터는 무엇입니까?

11. 귀 기관에서 보유하고 있는 데이터 중 외부 공개가 되지 않는다면 사유는 무엇입니까? (복수 선택 가능) (___)

① 보안 문제로 인한 공개 어려움
② 공개할 수 있는 물리적 방법(ICT시스템 등)이 없음
③ 공개 필요성이 없음 (요구하는 사람 없음)
④ 공개할 수 있는 절차나 규정이 조직 내에 정의되어 있지 않음

⑤ 데이터를 공개해도 보상이 전혀 없음
⑥ 기타 [ ]

12. 귀 기관에서 연구데이터 관리 측면에서 어려운 점(애로사항)이나 정부 등에 건의할 사항이 있다면 자유롭게 기술해 주시기 바랍니다.