

생명공학분야의 연구데이터 공유 사례에 관한 연구

A Study on Use Case of Research Data Sharing in Biotechnology

박 미 영 (Miyoung Park)*

안 인 자 (Inja Ahn)**

김 준 모 (Junmo Kim)***

초 록

연구데이터 관리·공유의 연구 패러다임 변화로 미국 및 유럽은 연구데이터 관리 의무화를 법제도화 하고 있으나 국내 연구현장은 인식부족과 인프라 미비로 연구데이터 관리지침 및 관리계획 등이 부재한 상황이다. 본 연구에서는 주요국가의 연구데이터 관리지침(미국 NISO DMP, 영국 UK 아카이브 DMP 등)을 조사하여 연구데이터 관리계획의 주요 단계를 도출하였다. 도출한 결과는 연구데이터 정책 계획 지원, 연구데이터 기술적 지원, 연구데이터 공유 활용 지원, 연구데이터 법적 매커니즘 지원, 연구데이터 교육 지원이다. 데이터 활용도가 가장 높은 생명공학분야 국내외 7개 기관의 연구데이터 공유 활용 사례를 연구데이터 관리 구성요소, 연구데이터 공유방법에 따라 비교 분석하였다. 유럽생물정보학연구소와 미국 국립생명공학정보센터는 연구데이터 공유 활용을 하고자 연구데이터 관리계획, 문서화, 데이터포맷, 데이터 저장, 공유 및 접근, 보존 등 관리를 위한 각 단계를 시행하고 있으나 국내의 경우 연구성과물 중 생물자원에 대한 제출만 이루어지고 있다. 연구데이터 관리계획(DMP)지원, 기술적 지원, 공유 활용 체계 지원, 법적 저작권 지원 등에 대한 가이드 및 매뉴얼 제작 배포로 국내 연구데이터 관리 및 공유 활용 대책을 연구데이터 관리계획 단계별로 마련해야 한다.

ABSTRACT

In this study, major steps of the research data management plan were derived through the research data management guidelines of major countries (NISO DMP in the US, UK DMP in UK archives, etc.). The results obtained are support for research data policy & planning, research data technical support, research data sharing support, research data legal mechanism support, and research data education support. In this study, we analyzed seven cases of data sharing among 7 domestic and foreign biotechnology. Shared use case countries are limited to the United Kingdom and the United States, which play a leading role in the management of research data. In Korea, shared cases were analyzed for the Korean Bio Information Center and related systems, which is a research and performance management and distribution agency designated by the Ministry of Science and Technology.

키워드: 연구데이터 공유, NISO, UK 아카이브, 생명공학

Research Data Sharing, NISO, UK Archives, Biotechnology

* Biotechnology Innovation Center, KSPP 수석연구원(mypark@kspp.re.kr) (제1저자)

** 동원대학교 문헌정보데이터관리과 교수(ijahn@tw.ac.kr) (교신저자)

*** Biotechnology Innovation Center, KSPP 연구원(jmkim@kspp.re.kr) (공동저자)

논문접수일자 : 2018년 2월 26일 논문심사일자 : 2018년 3월 15일 게재확정일자 : 2018년 3월 26일
한국비블리아학회지, 29(1): 393-416, 2018. [http://dx.doi.org/10.14699/kbiblia.2018.29.1.393]

1. 서론

1.1 연구의 배경 및 필요성

지난 20년간 학술연구방법이 크게 변화되었다. 새로운 소프트웨어, 하드웨어, 기계 및 대형 실험 장비들이 발명되고, 새로운 데이터 소스와 연구정보 또한 인터넷을 통해 전 세계에 확산됨으로써 전 세계의 연구자들이 전례 없는 속도로 연구내용과 방법 면에서 진전을 이루었다. 하지만 연구 패러다임의 변화는 새로운 문제들을 부각시켰는데 특히 연구결과의 재현성, 수행방법과 절차의 투명성 확보와 같은 문제들이다.

21세기에 연구를 수행하는데 있어서 이러한 당면 과제들을 해결하기 위해서는 올바른 연구 데이터 관리가 필수적이다. 연구데이터 관리계획을 신중하게 수립하여 문서화하고 데이터를 보존함으로써 재현가능하고 투명한 연구데이터를 얻을 수 있다.

연구데이터는 과학적 연구에서 사용된 일차 자료로서 연구자에 의해 직접 작성된 수치, 문자, 이미지, 음성 등의 사실적 기록이다. 연구데이터는 많은 시간과 비용을 들여 생성되는 귀중한 자원이며, 미래의 과학과 교육을 위해 사용되거나 재활용될 수 있는 인류의 자산이다. 연구데이터의 역할은 첫째, 연구자체에 대한

증빙과 검증 가능성, 둘째, 데이터의 중복수집 방지와 결과 확산, 셋째, 연구의 혁신과 잠재적인 새로운 데이터의 이용 촉진의 세 가지로 크게 요약될 수 있다(UK Data Archive 2011). 연구데이터의 공개와 공유는 연구과정과 결과에 대한 투명성과 책임을 극대화하고, 데이터를 통한 이용자와 생산자의 새로운 협력관계와 연구혁신을 이끌어 낼 수 있는 잠재력을 가지고 있다는 면에서 더욱 의미를 둘 수 있다.

최근 미국과 유럽을 중심으로 일정 규모 이상의 공적 연구자금을 지원 받은 연구과제에 대해 연구 과정에서 생성된 데이터의 관리를 의무화하는 규정들과 자료를 공유하기 위한 정책이 만들어지고 있다(UK Data Archive 2011, 6).¹⁾ 2010년 전후에 일어나기 시작한 연구데이터 관리 및 활용에 관한 가치인정과 요구는 시대적 흐름이며, 연구자금 지원기관으로부터 더욱 지지되고 있다. 지지하는 대표적 분야로 자연과학분야에 바이오테크놀러지 & 생물과학연구위원회(BBSRC, Biotechnology and Biological Sciences Research Council), 자연환경연구위원회(NERC, Natural Environment Research Council)를 들 수 있다. 특히 생명공학 분야의 데이터관리는 개인 연구자들이 확보하기 어려운 소재를 발굴, 수집, 보관하며, 연구자들에게 분양, 자문, 기탁 서비스를 제공하는 데에 가치와 의미가 있다.

1) 미국의 경우 미국국립과학재단(NSF, National Science Foundation)이 연구지원신청서에 데이터관리계획(DMP, Data Management Plan)을 포함하도록 요구하고 있다. 영국에서는 영국연구위원회(RCUK, joint Research Councils UK)가 연구데이터에 관한 성명을 준비 중이다. 공학물리과학연구회(EPSC, Engineering and Physical Sciences Research Council)는 연구데이터의 관리와 접근에 관한 정책을 수립하고 있으며, 의학연구회(MRC, Medical Research Council)는 연구데이터의 관리와 공유를 규정하는 보다 포괄적인 새로운 정책을 개발 중에 있다. 경제사회연구회(ESRC, Economic and Social Research Council)는 오랫동안 사용해 왔던 연구데이터의 정책과 지침을 개정했다.

국내에서는 실험·관측·분석 등 연구개발 과정에서 산출된 연구데이터는 다양한 활용가치를 보유하고 있으나 연구성과물로 포함되어 있지 않고 관리 대상에서 제외되고 있다. 데이터 관리 현황을 보면 일부 연구 분야에서 연구데이터 수집·관리 중이나 활용은 저조한 수준이다. 국내 소재정보은행에서 상업적으로 유통 중인 소재물성 DB를 보유하고 있으며, 국가참조표준센터에서 소재 물성 참조표준을 수집 및 제공하고 있는 수준이다. 한국과학기술정보연구원(KISTI)에 따르면 연구자의 82%가 연구데이터를 개인 자산으로 간주하고 개인적으로 관리하고 있으며, 관리·공유에 대한 연구현장 인식이 저조한 상황이다. 연구데이터 인프라 구축 현황은 공공기관 중심으로 연구데이터관리 서비스가 일부 진행되고 있으나 본격적으로 활성화되지 않은 실정이다. 한국과학기술정보연구원에서 과학데이터 공유와 보존을 위한 시스템 P-CUBE(Platform for Convergence research and Unification of Big E-resources)를 운영 중에 있으며 한국연구재단 기초학문자료센터(KRM)에서 인문사회분야 연구 자료를 수집 중에 있다. 또한 최근 2017년 한국정보통신기술협회(TTA)에서는 정보통신단체표준으로 연구데이터의 관리와 공유를 위해 메타데이터를 개발하였다.²⁾

데이터가 관리, 공유되기 위해서는 데이터의 품질 점검, 저작권, 개인정보, 장기보존 등 사전에 점검해야 할 사항들이 많다. 따라서 국가 차원에서 또는 소속 대학과 연구소 차원에서 교수와 연구자들의 연구 활동에 따른 데이터 관리를

다양한 측면에서 지원해야 할 필요성이 높아지고 있다. 디지털 데이터의 저장, 유통, 그리고 온라인으로 사용자가 접근할 수 있도록 하는 것은 어렵지 않기 때문에 많은 기관이 연구데이터를 공유함으로써 각자의 연구 영향력과 인지도를 높이기를 희망하고 있다. 반면에서 생명공학분야의 연구데이터는 개인차원에서 확보가 쉽지 않으며, 과학적 연구데이터의 관리와 공유를 통하여 재사용과 연구자들에게 분양, 자문, 기탁 서비스를 제공하는 데에 의의가 가장 두드러진 분야라고 할 수 있다. 그러므로 생명공학분야의 선도적 연구데이터관리는 미래 데이터관리에 영향을 미치는 중요한 활동이라고 할 수 있다. 따라서 본 연구에서는 국내외 연구기관에서 생성되는 생명공학 분야 연구데이터 관리 및 활용 현황과 이와 관련하여 국제적으로 운용되는 연구데이터 관리 표준 혹은 지침은 무엇인지 조사 및 분석하여 미래 연구데이터 관리 방안을 제시하는 것이 목적이다. 이를 통하여 국내 연구데이터의 공유 활용을 위한 기초연구에 기여하고자 한다.

1.2 연구의 목적

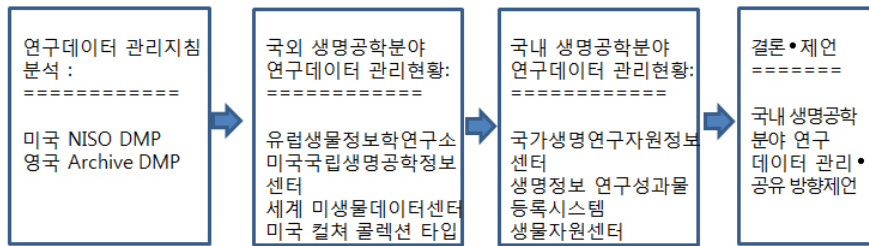
본 연구를 통하여 제시하는 연구는 문제는 다음과 같다.

첫째, 국내외 연구기관의 생명공학 분야 데이터는 어떻게 관리되고 있는가?

둘째, 국제 기준의 연구데이터 관리방안은 무엇인가?

셋째, 미래 국내연구기관의 생명공학 분야 데

2) 한국정보통신기술협회, 2017. 『연구데이터 관리 및 공유를 위한 메타데이터 정보통신단체표준(국문 표준, TTA, KO-10.0976)』.



〈그림 1〉 연구절차

이터는 어떻게 공유되어야 하는가?

연구 목적은 아래와 같다.

주요 국가의 연구데이터 관리지침을 통하여 연구데이터 관리계획의 주요 단계를 도출하고, 단계별로 국내외 생명공학분야의 연구데이터 공유 사례를 분석하였다.

이를 위하여 본 연구는 미국과 영국의 연구데이터 관리 구성요소의 비교분석으로 연구데이터 관리 표준 및 지침을 파악하였으며, 생명공학분야 해외의 4개 기관 및 국내 3개 기관의 연구데이터 공유 활용 사례를 조사하여 생명공학분야 연구데이터 관리 방안을 제시하고자 하였다 (〈그림 1〉 참조).

생성되거나 작성된 모든 정보로서 최종결과물 뿐만 아니라 중간생성물도 포함하며, 실험·관찰·시뮬레이션 데이터뿐만 아니라 이미지, 텍스트 등 다양한 형태의 정보도 포괄한다.³⁾⁴⁾ 연구데이터 유형에는 연구데이터는 원자료(raw data), 추출데이터(derived data), 출판된 데이터(published data)의 세 가지 유형으로 구분하며 원자료는 실험이나 측정, 관찰의 결과로서 직접적으로 생성된 데이터를 의미하고, 추출데이터는 데이터의 양을 줄이거나 물리적으로 의미 있는 좌표계로 변환하는 과정에서 특정 표준이나 절차에 의거하여 생성된 데이터를 말한다. 출판된 데이터는 출판물에 게시되거나 인용된 데이터로서 연구 결과를 도출하는 근거로 사용된 데이터를 말한다.⁵⁾

2. 연구데이터 관리의 이론적 배경

2.2 이론적 배경: 연구데이터 관리

2.1 주요 개념

연구데이터(Research data)란 연구 수행 중

영국정보시스템합동위원회(JISC)는 연구자를 지원하는 것과, 올바른 데이터 관리와 공유를 촉진시키는 것이 영국의 고등교육과 연구의

3) 국가과학기술심의회 운영위원회, 2018, 『연구데이터 공유활용 전략(안)』.

4) 미국 국립아카데미, 2009.10, 『Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age』.

5) Science and Technologies Facilities Council, 2011, *Scientific data policy* : 2.

이익을 위해 우선 사항이 된다고 결론을 내리고 영국 에섹스 대학을 선두로 『데이터의 관리와 공유』 우수사례 소책자를 발간하였다. 에섹스 대학의 UK 데이터 아카이브(UK Data Archive)에서 『데이터의 관리와 공유』를 위한 연구데이터 관리(DMP) 요소⁶⁾는 <그림 2>와 같다. 데이터 공유 타당성과 효과성을 제시하여 연구데이터 관리계획의 필요성을 설명하고 있으며 데이터 문서화, 데이터포맷, 데이터저장 등과 같이 연구데이터 공유를 위한 기술적 요구사항과 윤리 및 동의, 저작권, 공유전략 등 연구데이터 공유를 위한 법제도 요구내용을 세부적으로 제시하였다.

1) 데이터공유 타당성: 데이터의 공유는 새로운 과학연구를 쉽게 수행할 수 있도록

하고 데이터의 중복 수집을 방지하며 교육과 훈련을 위한 풍부한 현실적인 자원을 제공하는 효과가 있으며, 이를 근거로 데이터공유 타당성을 조사함

- 2) 연구데이터관리계획(DMP): 지속 가능하고 공유할 수 있는 고품질의 데이터를 생성하는 사전계획⁷⁾을 세우는 것이 필요함. 생산될 데이터 유형, 데이터 기술을 위한 표준, 데이터 접근계획, 데이터 재이용과 관련된 방침 및 데이터 보존계획을 기술한 문서가 포함됨
- 3) 데이터문서화: 데이터 파일 정리와 데이터 기술(포맷)이 포함됨. 데이터가 어떻게 만들어지고 디지털화 되었는지, 데이터의 의미와 내용 및 구조, 데이터에 실시된 모든 작업을 설명하는 것임



<그림 2> 연구데이터 관리 단계

* 출처: 한국과학기술정보연구원. 2017. 『데이터의 관리와 공유』

6) University of Essex. 2011. *Managing and Sharing Data*.

7) 한국과학기술정보연구원. 2017. 『데이터의 관리와 공유』.

- 4) 데이터포맷: 데이터변환과 구조화, 데이터 관리 등 플랫폼을 위한 포맷이 필요함. 포맷 구성요소로서 메타데이터는 수집 데이터의 기원, 목적, 해당 시기, 지리적 위치, 작성자, 접근 조건, 이용 조건을 설명하는 가장 중요하고 표준화, 구조화된 데이터 관련 문서의 일부임. 메타데이터는 검색가능정보 제공, 인용을 위한 서지 정보로 사용됨
- 5) 데이터저장: 데이터의 저장 및 백업 관련 서비스로서 데이터저장, 보완, 전송 및 암호화 및 공유와 협력사항이 포함됨
- 6) 윤리 및 동의: 법적 및 윤리적 문제 설명과 이해에 근거한 동의가 포함됨
- 7) 저작권: 데이터와 관련된 저작권 및 라이선싱을 안내함
- 8) 공유전략: 연구센터와 연구프로그램은 가장 적절하고 실천적인 데이터 관리 프레임워크를 제공해야 함

2.3 선행 연구

연구데이터의 법, 제도에 관한 연구, 연구데이터 관리현황연구, 레포지터리 연구, 메타데이터에 관한 연구로 구분할 수 있다. 가장 많은 연구는 관리현황에 대한 연구로서 관리에 대한 일반적 필요성 및 메타데이터 필요성 등이 제시되고 있다. 반면 실제적인 관리체계를 위한 영역별 구분 및 관리내용에 관한 연구는 매우 부족한 것으로 조사되었다. 법, 제도 및 관리정책에 대한 연구는 외국 선진사례를 중심으로 기준을 제시하여 세부내용을 기술되었다. 연구데이터 공유를 전제로 하는 레포지터리 활성화 연구가 다수 시행되었으며, 공통적 정책요소를 제시하였다. 메타데이터에 관한 요소 연구가 시행되었으며, 메타데이터 표준, 웹서비스 표준 기술, 학술정보서비스그룹, 데이터 관계기술그룹으로 구분하여 요소도출방법을 사용한 것이 특징이다(〈표 1〉 참조).

〈표 1〉 선행연구 요약

	상세
법 제도	- 김규민(2015) 과학데이터관리에 관한 해외 법제의 비교법적 고찰 - 윤종민, 김규빈(2013) 과학데이터에 관한 입법례와 관리정책 그리고 대응방안
관리정책	- 김지현(2013) 국외 정부연구비지원기관의 연구데이터 관리정책 분석
연구데이터 관리	- 강주연(2017) 생명공학분야 연구데이터 관리 방안 연구 - 김지현(2015) 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구 - 김지현(2012) 대학 내 연구자들의 연구데이터 관리에 관한 연구 - 김지현(2011) 자연과학분야 대학실험실에서의 연구노트 작성 및 관리에 관한 연구
레포지터리	- 김문정(2015) 과학기술분야 연구자의 레포지터리 활성화를 위한 연구데이터 공유에 대한 연구 - 김지현(2016) 연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구
연구정보 수집	- 김은정, 남태우(2012) 연구데이터 수집에 영향을 미치는 요인 분석
메타데이터	- 김선태(2012) 해양관측분야 메타데이터 표준요소 선정 및 설계에 관한 연구 - 김선태, 한선화, 이태영, 김용(2010) 과학데이터 보존 및 활용모델에 관한 연구.

2.3.1 법 제도에 관한 연구

윤종민, 김규빈(2013)은 “과학데이터에 관한 입법례와 관리정책 그리고 대응방안 - 호주, 미국, 중국을 중심으로”, 김규빈(2015)은 “과학데이터관리에 관한 해외 법제의 비교법적 고찰”을 통하여 호주, 미국, 중국 등은 국가차원에서 관련 기관을 통해 과학데이터를 수집, 관리 및 유지하는 등 데이터 활용에 관하여 조사 분석하였다. 이들 나라들은 과학데이터 활용을 위하여 중장기적인 정책수립, 법제도 정비, 기반시설에 대한 투자를 지속적으로 확대하고 있다. 각 나라별로 과학데이터 관리지침과 관리체계를 제시하였다.

2.3.2 관리정책에 관한 연구

김지현(2013)은 “국의 정부연구비지원기관의 연구데이터 관리정책 분석”을 통하여 미국, 영국, 캐나다 등 15개 정부연구비지원기관의 데이터관리정책을 분석하였다. 정책내용을 데이터 정의, 관리원칙, 관리계획, 관리실행, 법적·윤리적 측면의 5가지 기준에서 분석한 결과 모든 기준을 포함한 데이터정책은 없지만 다수의 기관이 분석기준내용을 공통적으로 다루고 있다. 이를 통하여 국내 데이터정책개발 제언을 하였다.

2.3.3 연구데이터 관리에 관한 연구

김지현(2011)은 “자연과학분야 대학실험실에서의 연구노트 작성 및 관리에 관한 연구”에서 물리, 화학, 생명과학 실험실의 교수들과 대학원생들을 대상으로 반구조화된 심층 면담을 실시하여 연구노트 작성 현황을 분석하였다. 결과 연구노트의 중요성을 인식하지만 형식보다

는 내용적 측면에 초점을 맞추었으며, 견본노트를 통하여 작성법을 익히고 있다. 논문에서는 연구노트 생산맥락을 파악함으로써 대학기록관의 역할을 강조하고 있다. 다음 해 김지현(2012)은 “대학 내 연구자들의 연구데이터 관리에 관한 연구”를 통하여 한국연구재단 지원사업의 최종선정과제 대학 소속연구자를 중심으로 연구데이터 관리방식과 관련하여 본 연구에서는 데이터의 가치를 높이는 처리 방식과 데이터보관 장소 및 데이터의 보존 연한에 대한 인식을 조사하였다. 연구데이터의 관리체계가 존재하지 않는 경우가 대부분이었으며 특히 데이터의 장기보존과 메타데이터 작성에 대한 만족도가 낮아 이를 지원할 수 있는 연구데이터 아카이빙 서비스의 개발이 요청된다. 김지현(2015)에는 “데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구”를 통하여 국내 대학에 소속된 연구자들을 대상으로 데이터 관리 현황과 데이터 공유 및 재이용에 대한 경험과 인식을 조사하였다. 데이터를 보존하려는 연구자들이 많아 데이터의 저장과 보존에 대한 연구자들의 요구가 적지 않은 것으로 나타났다. 데이터의 공유와 재이용은 개인적인 연구 모임이나 연구 팀 등 잘 알고 있는 사람들과의 범위 내에서 이루어지고 있었다. 다수의 참여자들이 연구아이디어의 도용, 표절, 논문 출판의 주도권 문제 등 연구데이터 공유에 대한 우려를 나타내고 있었으며 이를 완화할 수 있는 유인책이 마련될 필요가 있다.

강주연(2017)은 “생명공학분야 연구데이터 관리 방안 연구”에서 우수사례를 중심으로 주제 전문분야의 관리현황을 분석한 자료이다. 해당분야의 연구데이터의 관리, 정책, 보존, 공유와

관련한 연구의 동향을 살펴보았다. 특히 연구데이터 관리 현황 분석을 위한 도구로 DAF(Data Asset Framework)를 개발하여 관리를 위한 연구특성, 기초정보, 데이터기술, 저장, 보존, 접근, 공개 등의 관리요소를 추출하여 동향을 분석하였다.

2.3.4 레포지터리에 관한 연구

김문정, 김성희(2015)는 “과학기술분야 연구자의 레포지터리 활성화를 위한 연구데이터 공유에 대한 연구”에서 오픈액세스 환경에서의 레포지터리 공유에 미치는 영향요인을 분석하였다. 매개변수인 인지성만이 보상체계를 통하여 연구데이터 공유에 긍정적인 영향을 미치는 것으로 나타났으며 신뢰성을 제외한 의사소통의 개방성, 협력성 또한 영향을 미치지 않는 것으로 나타났다. 또한 보상체계와 연구데이터 공유에 통계적으로 유의미하게 영향을 미치는 것으로 나타났다. 김지현(2016)은 “연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구”에서 해외 연구데이터 레포지터리 37곳을 대상으로 데이터 접근 및 이용 통제를 규정하고 있는 정책 요소를 분석하였다. 자연과학 30분야, 사회과학 및 일반과학 분야 7개 레포지터리로 구분하여 분석을 실시한 결과 저작권 및 라이선스 규정, 데이터 인용, 면책조항 및 엠바고 적용 관련 규정이 공통적으로 제시되는 정책 요소인 것으로 나타났다. 그러나 분야별로 규정되고 있는 정책 요소의 다양성에는 차이가 있는 것으로 나타났다.

2.3.5 연구정보수집에 관한 연구

김은정, 남태우(2012)는 “연구데이터 수집에

영향을 미치는 요인 분석”에서 학술과제관리규칙, 연구노트 지침 등 요인항목을 도출하여 분석한 결과 내적 동기부여요인, 정보인프라요인, 연구데이터 제출에 장애요인 최소화를 통하여 수집활성화 방안을 제시하였다.

2.3.6 메타데이터에 관한 연구

김선태, 한선화, 이태영, 김용(2010)은 “과학데이터 보존 및 활용모델에 관한 연구”에서 과학데이터와 관련하여 미국과 영국, 호주, 유럽연합의 국가별 동향을 조사하였으며, 해외 과학데이터 서비스 선진 사례를 프로그램 단위로 조사하였다. 과학데이터의 관리 및 활용을 위해서는 이질적 과학데이터에 대한 통합 메타데이터 스키마의 개발에 관한 연구가 필요하다.

김선태(2012)는 “해양관측분야 메타데이터 표준요소 선정 및 설계에 관한 연구”에서 해양관측분야 메타데이터 표준, 웹서비스 표준기술, 학술정보서비스그룹, 데이터 관계기술그룹으로 구분하여 Identifier, date 요소, Coverage 요소, CitationServingFlag, Format 요소, Recordtype 요소 등을 추출하였다. 이는 하이브리드 요소도출방법을 사용한 것이 특징이다.

이상의 선행연구를 통하여 우리나라의 연구데이터에 관한 개념과 인식이 이제 태동되는 시기이며 연구 데이터의 관리에 대한 필요성을 강조하는 논문이 몇 편 발표된 정도이다. 결론적으로 연구데이터 관리연구가 서비스 기술 분류에서 근거하여 세분화된 연구가 실시되기보다는 필요성, 인식, 연구데이터관리의 요구 등 시행되었기 때문에 데이터 파일 정리, 기술, 데이터 저장, 데이터 관리 교육 등의 선행연구 부족하다.

반면 국외의 경우는 지침서, 법제도의 기반 하에 데이터관리 계획 및 가이드라인 및 우수 사례가 제시되어 있다. 따라서 해당 부분에 대한 기능 필요하며 이에 대한 동향과 사례를 살펴보는 것이 필요하다. 따라서 본 논문에서는 연구데이터의 가치와 요구가 가장 강한 분야 중의 하나인 국내의 생명과학분야의 연구데이터에 관한 관리 및 서비스 동향을 살펴보고자 한다.

3. 국내외 연구데이터 관리 단계분석

연구데이터 관리의 선진적 역할을 하고 있는 미국과 영국에서는 연구관리기관을 중심으로 연구데이터 공유 및 활용을 위한 관리 지침서 및 매뉴얼을 발간하고 있다. 본 논문에서는 미국 정보표준화기구(National Information Standards Organization, NISO)의 연구데이터 관리에 있어서 필요한 구성요소, 영국 UK·데이터아카이브의 데이터관리와 공유의 주요 요소, 국내 연구데이터관리서비스 기술요소를 비교분석하고자 한다.

UK Data Archive(2011)는 연구데이터 관리와 공유 활동에 기초 지식을 제공하는 지침서로서 2009년 “Managing and Sharing Data: Best Practice For Researchers” 초판, 2011년 제3판을 발행하였다. 데이터공유의 이유와 방법, 관리계획, 데이터문서화, 포맷, 저장, 저작권 등의 데이터 공유를 위한 기본 틀을 제시하는데 중점이 있다. 2014년 일본어번역판(동경 대학 사회과학연구소 부속 사회조사·데이터

아카이브 연구센터), 2017년 한국어번역판(한국과학기술정보연구원)이 발행되고 있으며, 이를 통하여 연구데이터의 관리와 공유에 대한 필요성이 전 세계적으로 확산되어가고 있는 추세를 발견할 수 있다.

앞서 언급된 연구데이터 공유 및 활용을 위한 관리 지침서 및 매뉴얼은 데이터 관리와 공유 측면에서 공유 정책과 연구데이터관리 계획상의 주요 요소들을 다루고 있다. 데이터관리 계획(DMP), 연구데이터 문서화, 데이터 기술(포맷), 데이터 저장, 데이터 공유 및 접근(사용), 데이터 보존, 데이터 인용/출판, 데이터 관리 교육, 데이터 저작권의 9요소이다.

미국의 경우 2015년 정보표준화기구(National Information Standards Organization, NISO)에서 발간한 “Research data Management”는 연구데이터 사용자를 위한 입문서(Primer) 시리즈의 하나로서 최근의 연구데이터 발전과 새로운 도구, 모범사례 및 사용 가능한 리소스에 대해 기술하고 있다.

국내(김지현 2014)는 연구데이터관리 서비스의 구성요소 9가지를 기준으로 서비스의 현황과 내용을 분석하였다. 서비스의 구성요소 9가지로 (1) DMP 작성지원; (2) 데이터 파일 정리; (3) 데이터 기술; (4) 데이터 저장; (5) 데이터 공유 및 접근; (6) 데이터 보존; (7) 데이터 인용; (8) 데이터관리 교육; (9) 데이터 지적재산권을 포함한다.

Veerle Van den Eynden(2011)은 영국 UK 아카이브와 미국 정보표준화기구(NISO)의 연구데이터 관리항목을 <표 2>와 같이 제시하고 있으며, 앞서 도출된 국내 연구데이터관리 서비스 구성요소 9가지 기준과 비교분석하여 국

〈표 2〉 연구데이터관리 분석결과 종합분석

국내 연구데이터 관리 구성요소	영국 UK Archive 「연구데이터 관리 및 공유」	미국 NISO 「연구데이터관리」
데이터관리계획 (DMP)	역할과 책임 데이터관리의 비용계산	데이터 설명, 기준, 정책과 절차, 아카이빙과 보존, 필요한 자원 명명 스키마 설계, 스프레드시트 설계, 메타데이터 수집 계획 수립, 백업전략 수립, 데이터관리계획을 위한 인적자원
연구데이터 문서화	데이터의 문서화 메타데이터	연구데이터 문서화, 업무절차 문서화
데이터 기술(포맷)	파일포맷, 데이터 변환 파일과 폴더의 체계화, 품질보증, 버전관리와 신뢰성, 트랜스크립션	메타데이터, 문서화 소프트웨어, 업무절차 소프트웨어, 컴퓨터 시뮬레이션
데이터 저장	백업생성, 데이터의 저장 및 보안, 데이터의 전송과 암호화, 데이터의 폐기	데이터스토리지, 백업, 보안, 백업 버전통제, 클라우드 스토리지
데이터 공유 및 접근(사용)	파일공유와 공동 환경	데이터셋 식별과 연결, 식별자, 데이터인용과 관련된 이슈, 인용
데이터 보존	--	레포지토리, 분야별 레포지토리, 일반 레포지토리, 레포지토리 선택, 레포지토리 소프트웨어
데이터 인용/출판	--	데이터 출판(가용성, 인용성, 유효성), 데이터출판 모델, 보상과 인센티브
데이터 관리 교육	--	--
데이터 저작권	법적 및 윤리적 문제, 설명과 이해에 근거한 동의와 데이터 공유, 데이터의 익명화 접근 관리, 데이터의 재사용과 저작권	데이터셋 거버넌스와 사용 등의 저작권 및 소유권, 데이터베이스권 데이터관리기술을 위한 법적 메커니즘 중요 데이터

* 출처: Veerle Van den Eynden, Matthew Louise Corti, Libby Bishop Woollard, and Laurence Horton(2011).

내외 공통의 연구데이터 관리 구성요소를 도출하였다.

분석결과를 종합적으로 정리해 보면 〈표 2〉와 같이 연구데이터관리는 크게 1) 연구데이터관리 정책과 계획 수립 지원, 2) 연구데이터관리의 기술적 지원, 3) 연구데이터관리 교육 지원, 4) 연구데이터 공유 및 활용 지원, 5) 법적 윤리적 지원(저작권 및 지적재산권)으로 분류할 수 있으며, 데이터 보존, 인용, 출판, 관리교육 등은 연구데이터 관리 구성요소로서 일부만 제시하고 있다.

4. 생명공학분야 공유 사례 분석

연구 활동에서 만들어진 데이터는 미래의 과학과 교육을 위해 사용되거나 재활용될 수 있는 귀중한 자원이다. 데이터의 공유는 새로운 과학연구를 쉽게 수행할 수 있도록 하고, 데이터의 중복 수집을 방지하며 교육과 훈련을 위한 풍부한 현실적인 자원을 제공한다. 본 장에서는 국내외 생명공학분야 주요기관별 연구데이터 공유방법을 〈표 3〉과 같이 조사하였고, 연구데이터 공유방법별 연구데이터 관리 구성요소를 확인함으로써 국내외 생명공학분야 연구데이터 공유 및 활용 방안을 제시하였다.

〈표 3〉 해외 주요기관별 연구데이터 공유사례 비교

연구데이터 공유방법	기관 리파지토리 기탁	데이터 뱅크 기탁	전문데이터 센터 기탁	
주요기관	유럽 생물정보학연구소 (EBI)	미국 국립생명공학정보센터 (NCBI)	세계 미생물데이터센터 (WDCM)	미국 컬처 콜렉션 타입 (ATCC)

4.1 생명공학 자원

4.1.1 생명공학의 범주

생명공학(biotechnology)은 “생체나 생체유래물질 또는 생물학적 시스템을 이용하여 산업적으로 유용한 제품을 제조하거나 또는 공정을 개선하기 위한 기술”이다(한국생명공학연구원 홈페이지). 연구영역으로는 크게 인체, 동물, 미생물, 식물로 나눌 수 있으며 상기 4가지의 분류 안에서 세포, 염색체, 계놈, DNA, 유전자 등으로 세분화 될 수 있다. 생명공학은 전자 공학처럼 그 자체가 생산품이 아니며 그 대신 다른 산업 부분에서 중요하게 기여할 기술의 범위를 다루고 있다(오계현 외 2005). 이러한 이유로 다른 분야의 자연과학과 달리 다른 학문으로의 영향력이 확장되고 있다. 한국생명공학연구원에서는 이러한 관점에서 바이오 인포메틱스, 바이오 케미칼, 바이오 일렉트로닉스, 바이오 리메디에이션, 바이오 의약 유전자 치료, 바이오 메커닉스, 고기능 작물, 기능성 식품에 대한 연구를 생명공학 분야의 응용연구 분야로 제시하고 있다.

본 연구에서는 생명공학분야의 연구데이터 관리 및 공유 현황의 선진사례를 제시하고자 한다.

4.1.2 생명연구자원의 정의 및 분류

생명연구자원은 생명공학연구의 기반이 되는 자원으로 산업적으로 유용한 동물, 식물, 미생물, 인체유래 연구자원 등 생물체의 실물과 정보를 말한다(생명연구자원의 확보·관리 및 활용에 관한 법률 제2조 1항(2016.9.23. 시행)). 생명연구자원은 생물자원, 생물다양성, 생명정보를 총체적으로 포함하는 개념으로 배양 가능한 생물체(예, 미생물, 식물, 동물 및 인간 세포주), 이들의 복제 가능한 부분(예, 계놈, 플라스미드, 바이러스, cDNAs), 미배양 생물체, 세포주 및 조직, 생명자원의 분자적, 생리적, 구조적 정보에 대한 데이터베이스와 관련 정보들이다(OECD 모범운영지침 2007).

생명공학적인 면에서의 연구소재(research resource)는 과학기술의 기반을 이루는 연구재료로 생물자원과 비생물자원의 실물 및 관련 정보를 말한다. 그 중에 생물자원은 생물체(미생물, 식물, 동물, 바이러스 등), 이들의 파생물(조직, 세포, 핵산, 단백질, 추출물 등)과 인체 유래물⁸⁾이며 비생물자원은 광물질, 금속, 반도체, 합성화합물 등이다.

이와 관련하여 연구소재 은행(바이오 뱅크)에서는 개인 연구자들이 확보하기 어려운 인체 유래물, 동물, 식물, 미생물, 융합물질 등 다양한 분야의 소재를 수집해 특성을 규명, 보관하고 연구

8) 인체유래물은 인체로부터 수집하거나 채취한 조직, 세포, 혈액, 체액 등의 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, 디옥시리보핵산(DNA, deoxyribonucleic acid), 리보핵산(RNA, ribonucleic acid), 단백질 등 임.

구자들에게 분양, 자문, 기탁 서비스 제공한다.)⁹⁾

4.2 국내외 공유사례 분석

경제협력개발기구(OECD)의 지침에 따르면 공적인 연구자금 제공자는 공적 자금을 지원받은 연구에서 생성된 데이터를 가능한 한 과학 커뮤니티에 공개·이용되도록 하여야 한다. 따라서 많은 연구자금 제공자는 연구데이터 공유 정책을 채택하고 있으며 연구자에게 데이터와 결과의 공유를 의무화 또는 권유하고 있다.

연구데이터를 공유하는 방법에는 다양한 방법들이 있다. 1) 전문데이터센터, 데이터 아카이브, 데이터뱅크에 기탁, 2) 출판물의 보조자료로서 학술지에 제출, 3) 기관 리포지토리에 기탁, 4) 프로젝트 또는 기관의 웹 사이트를 통해 온라인으로 이용할 수 있는 방법, 5) 비공식적으로 연구동료(peer to peer) 평가기반의 연구자들 사이에서 공유하는 방법이 있다(UK Archive 2017). 본 장에서는 공유 방법에 따른 생명공학 분야에서의 사례를 살펴보면 아래와 같다.

4.2.1 국외 사례

해외 주요기관별 연구데이터 공유 방법으로는 기관 리포지토리 기탁방법의 유럽생물정보학연구소(EBI, European Bioinformatics Institute), 데이터뱅크 기탁방법의 미국 국립생명공학정보센터(NCBI, National Center for Biotechnology Information), 전문데이터센터 기탁 방법은 세계

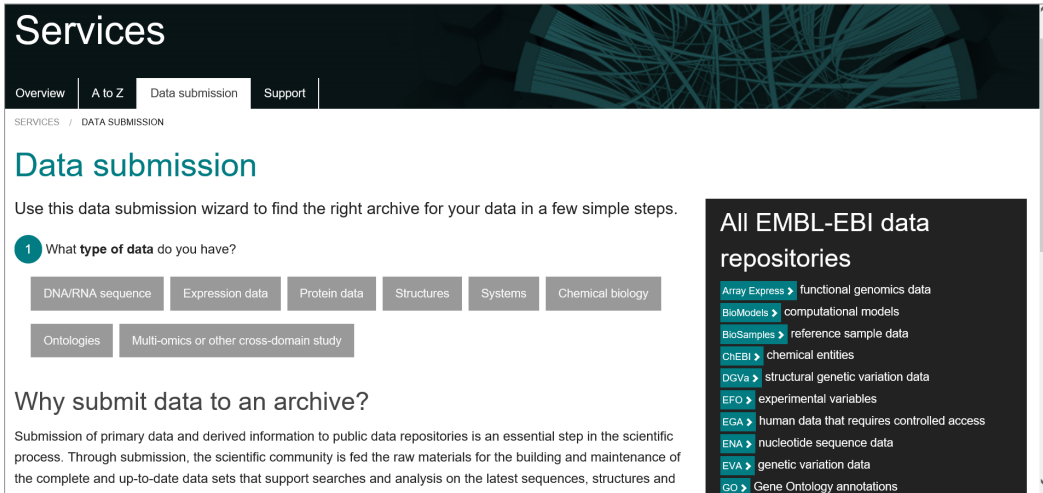
미생물데이터센터(WDCM, World Data Centre for Microorganisms), 미국 컬처 콜렉션 타입(ATCC, American Type Culture Collection) 사례를 다음과 같이 분석하였다.

- 1) 기관 리포지토리 기탁: 유럽생물정보학연구소(European Bioinformatics Institute) 기관 사례 바이오 분야 EMBL-EBI

연구데이터 리포지터리는 데이터를 장기적이고 신뢰할 수 있는 방식으로 저장하고 접근할 수 있게 하는 정보 인프라를 의미한다. 여기에는 다양한 형태의 데이터 아카이브와 데이터센터 또는 디지털 도서관이 포함된다(Pompel et al. 2013). 영국의 Hinxton에 위치한 유럽생물정보학연구소(EBI)는 유럽분자생물학실험실(EMBL)의 산하 기관으로 1974년 설립된 유럽의 대표적인 생물정보학 연구기관이다. 유럽생물정보학연구소(EBI, European Bioinformatics Institute)의 바이오모델(bio-model) 데이터베이스는 생물학 프로세스와 분자 기능에 대해 동료 평가를 거쳐 출판되고 있는 리포지토리이다. 모든 모델에는 주석이 붙어있고, 관련된 데이터 리소스에 연결된다. 연구자들에게는 오픈소스 포맷인 Systems Biology Markup Language (SBML)로 쓴 모델을 기탁하도록 권유하고 있으며 모델은 장기적인 보존용으로 정리된다.

유럽생물정보학연구소(EBI)에서는 1차 데이터와 파생 정보를 공개 데이터 저장소에 제

9) Biological Resource Centers(BRC): 생명공학의 필수기반으로서 살아있는 세포, 계능, 생명기능 및 유전에 관련된 정보를 보존하고 제공하며 과학, 산업, 농업, 환경, 의료 연구개발과 응용에 필요한 생물자원의 보존과 분양, 생물자원에 대한 연구개발 수행, 생물다양성 보전, 생물 레퍼런스 물질에 대한 보존, 지식재산권 보호를 위한 생물자원의 보존, 공공정보와 정책수립에 대한 기초자료 제공 등의 기능을 수행하는 곳을 말함(OECD 모범 운영지침 2007).



〈그림 3〉 EMBL-EBI 연구데이터 리포지토리

〈표 4〉 EMBL-EBI 연구데이터 유형 및 리포지토리(데이터저장소)

데이터 유형	분석도구	데이터 저장소	데이터베이스
<ul style="list-style-type: none"> • DNA/RNA sequence • Expression data • Protein data • structures • systems • Chemical biology • Ontologies • Multi-omics or other cross-domain study 	<ul style="list-style-type: none"> • Clustr Search • WSDbfetch • WSNBBI Blast • WScLustalW • transeq • pepwindow • getorf • SRS • MSDfold • DALI • Expression Profiler • EMBL Tools 	<ul style="list-style-type: none"> • Array Express • Biomodels • BioSamples • ChEBI: chemical Entities • DGVa: Structural genetic variation data • wwPDB: Electron microscopy, X-ray data • Crystallography & NMR data 	<ul style="list-style-type: none"> • UniProt • UniPrptKB/SwissProt • EMBL • GOA • InterProScan • chEBI • chSTr

출하는 것은 과학적 프로세스에서 필수적인 단계로 간주하여 최신 시퀀스, 구조 및 분자 프로파일에 대한 검색 및 분석을 가능하게 하는 데이터 세트의 작성 및 유지를 위한 서비스를 제공한다. 또한 다양한 유형 및 데이터 규모에 적합한 서비스를 연구자들에게 제공한다. 주요 서비스를 보면 아래와 같다.

- 데이터베이스: 생물정보학 분야에 관한 순수 및 응용 연구, 핵산, 단백질 서열 등의 생물학적 데이터 구축·관리·제공

- 산업체 지원: 생명정보 데이터의 저장 및 개발 지원, 생명공학, 화학, 제약 기업 등의
- 분석도구: 생명정보 데이터베이스를 분석할 수 있는 다양한 도구를 제공

또한 〈표 5〉와 같이 다양한 분석도구를 제공하고 있다.

연구데이터 레포지토리 입장에서는 데이터 접근 및 이용을 통제함으로써 데이터의 무결성(integrity)을 강화할 수 있고 이를 통해 레포지

〈표 5〉 EMBL-EBI 분석도구

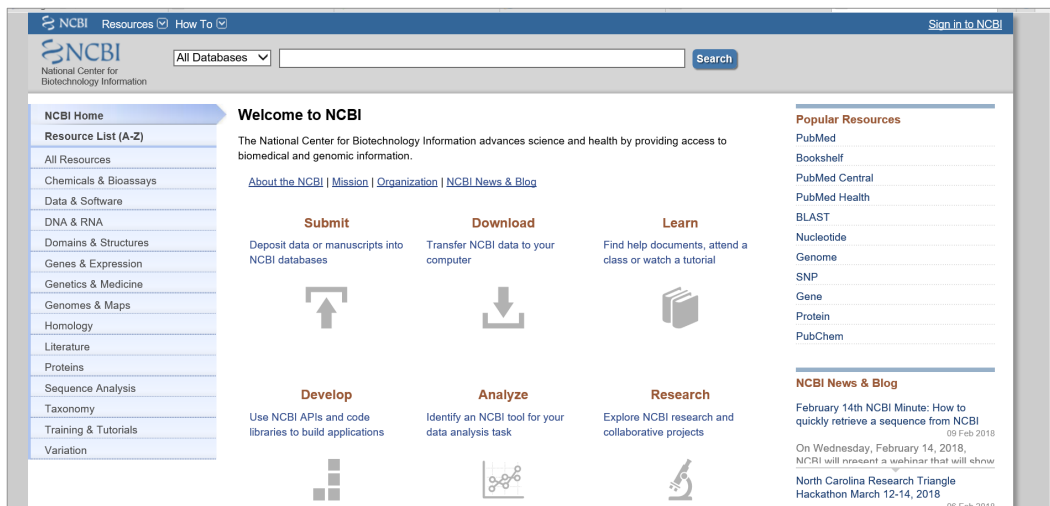
콘텐츠	설명
WSDbfetch	- EBI의 대표적인 웹서비스를 통해 제공되는 도구 - 다양한 생물학 데이터베이스를 자신이 원하는 포맷으로 검색 - 데이터베이스에 따라 fasta, xml 등의 데이터베이스 고유 포맷 및 raw 및 html형태의 스타일을 지원
WSFasta	임의의 서열과 유사성을 가진 서열을 서열 데이터베이스로부터 검색
WSNCBI Blast	NCBI에서 제공되는 BLAST2(Basic Local Alignment Search Tool)
WSLustalW	핵산이나 단백질의 전역다중정렬(global multiple alignment)에 사용되는 자동화된 프로그램
Transeq	핵산서열을 단백질 서열로 변환
pepwindow	단백질의 hydropathy를 보여줌
getorf	서열로부터 open reading frames(ORFs)를 찾음

터리의 지속가능성과 데이터의 기탁 비율을 높일 수 있다는 주장도 제기되고 있다(Eschenfelder and Johnson 2014). 반면 기관 리파지토리는 데이터의 장기적인 관리나 매우 복잡한 연구데이터에 대해 지원하지 못할 가능성이 있다. 웹사이트는 대부분 지속 가능성이 거의 없는 단기적인 것이다.

2) 데이터뱅크에 기탁: 미국 국립생명공학정보센터(National Center for Biotechnology

Information) 데이터뱅크는 데이터를 신뢰할 수 있는 방식으로 저장하고 접근할 수 있는 데이터베이스이다. 미국 국립생명공학정보센터(NCBI) 미국국립보건원 산하로 1988년 설립되었으며 전 세계 생명과학자들이 가장 많이 이용하는 사이트이다.

- 데이터베이스: 염기서열 데이터베이스인 GenBank를 운영하며, 데이터 분석과 검색 서비스를 제공한다. 연구자가 BioProject



〈그림 4〉 NCBI 데이터 기탁 서비스

<표 6> NCBI 연구데이터 관련 서비스

소개	미국립보건원(NIH: National Institutes of Health) 산하로 1988년 설립
활동	- 전세계 생명과학자들이 가장 많이 사용하는 사이트 - 염기 서열 데이터베이스인 Genbank 운영 - 데이터 분석과 검색 자원을 제공
주요 콘텐츠	- Conserved Domains : 단백질 도메인의 데이터베이스 - Genome 데이터베이스 : 다양한 계통, 완전 염색체, 콘택과의 서열 지도 및 통합 유전 및 물리 지도에 대한 보기를 제공 - Nucleotide : 서열 데이터베이스의 3자 구성 국제 공동 작업의 회원인 GenBank, EMBL 및 DDBJ의 서열 데이터가 포함 - OIMM 데이터베이스 : 인간 유전자 및 유전병 데이터베이스 - PopSet 데이터베이스 : 모집단, 계통발생 또는 돌연변이 연구의 결과 세트에 제출되는 정렬된 서열이 포함 - ProbeSet : 시약 배포자, 탐침 유효성 및 계산된 서열 유사성에 대한 정보와 함께 광범위한 생물학 연구 응용 프로그램에서 사용하도록 설계된 핵산 시약의 공개 레지스트리 - Structure/ MMDB(Molecular Modeling Database) : 결정학적 및 NMR 구조 분석의 실험 데이터가 포함, MMDB의 데이터는 PDB(Protein Data Bank)에서 가져옴

인터페이스를 통해 유전체 및 유전자 데이터를 제출하고 데이터 유형에 맞는 소프트웨어를 사용할 수 있다. GenBank 데이터베이스에 연구데이터 제출 프로세스가 제공되며 웹 기반 시퀀스 제출 도구도 제공하고 있다.

- 연구데이터 제출: GenBank, EMBL, DDBJ에 제출한 후 업데이트하기 위해 미국 국립생명공학정보센터에서 개발한 소프트웨어도 제공하며 짧은 mRNA 서열, 긴 서열을 포함하는 복잡한 제출물, 다중 주석, 분절된 DNA 세트 등을 처리 할 수 있다. 간단한 제출을 위해서는 BankIt 대신 온라인 제출 도구를 사용할 수도 있다.

3) 전문 데이터 센터에 기탁: 세계 미생물 데이터센터(WDCM, World Data Centre for Microorganisms): 중국 과학원 미생물 연구소(IMCAS) 산하기관

데이터 센터는 저장하는 데이터를 평가하고

선정하기 위해 특정한 기준을 적용한다. 전문적인 데이터 센터에 데이터를 기탁하는 경우 표준화된 접근 가능한 데이터 포맷으로 데이터를 장기적으로 저장한다. 소프트웨어의 업그레이드나 변경에 따라 필요한 경우 포맷의 변환이 이루어진다.

가장 대표적인 전문적인 데이터 센터인 WFCC (World Federation for Culture Collections)는 전 세계의 컬처 컬렉션에 대한 포괄적인 정보를 제공한다. WFCC의 전문가 중 호주 퀸즈랜드 대학의 V. B. D. Skerman 교수와 그의 동료들은 1960년대 전 세계의 컬처 자원에 대한 국제 데이터베이스 개발을 시작했다. 이와 관련하여 중국 과학원 미생물 연구소(IMCAS)는 2011년 WCCC-MIRCEN의 WDCM(세계 문화 수집연합 세계 미생물 데이터 센터)를 시작했다. WDCM은 중국에 정착한 생명 공학 분야의 최초 데이터 센터로서 중국의 미생물 연구를 넘어 세계 미생물 자원의 데이터 공유 플랫폼을 구축하고 있다. 발전하여 WFCC의 세계 미생물데이터 센터(WDCM)가 되었다.

WFCC INFO 데이터베이스의 레코드에는 조직의 조직, 관리, 서비스 및 과학적 이해 관계에 대한 데이터가 포함된다. 포함자원과 내용은 다음과 같다.

- 데이터베이스: 각 기록은 조류 수집, 시아노박테리아, 박테리아, 균류, 효모, 이끼류, 원생 동물, 조직 배양 및 바이러스와 같은 걸쳐 수집품의 목록인 STRAIN 데이터베이스에 저장된다.

전문 데이터센터에서는 카탈로그를 통해 데이터가 인터넷에서 검색대상 자원이 되고, 많은 이용자에게 데이터를 배포할 수 있으며 데이터의 재이용을 감독할 수 있다. 나아가 데이터 소유자를 대신하여 데이터에 대한 접근권을 관리하고 이용자의 질의에 대응한다는 면에서 공유활용의 가치는 있는 형태이다.

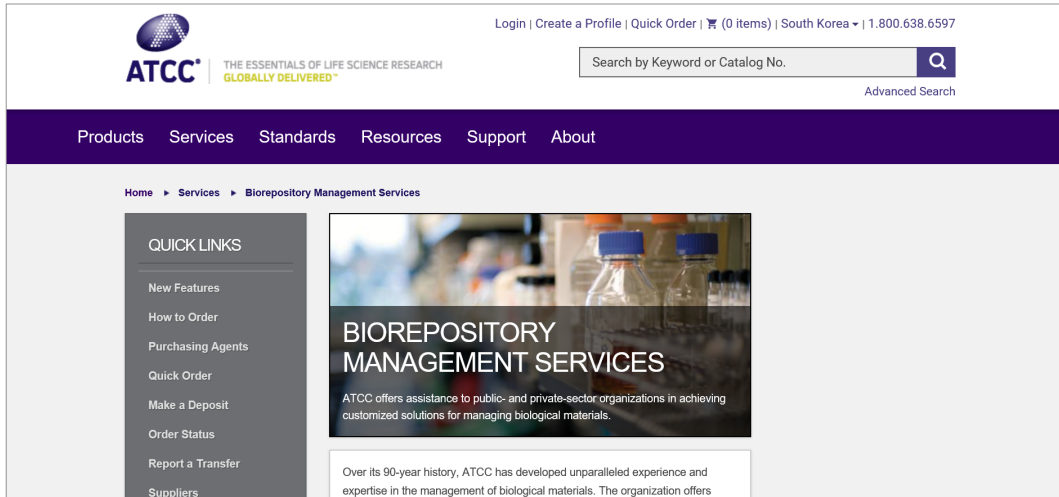
4) 데이터 아카이브에 기탁: 미국 걸쳐 컬렉션 타입(ATCC, American Type Culture Collection)

ATCC는 생물 자원 관리를 위한 맞춤형 솔루션을 공공 및 민간 부문에 서비스를 제공하며 세포주, 분자 유전체 도구, 미생물 및 생물 제품을 비롯한 광범위한 연구용 생물학 데이터 및 컬렉션을 보유하고 있으며 세부 내용은 아래와 같다.

- 생명연구자원정보관리: 세포주, 세균 균주, 바이러스, 곰팡이, 효모 등
- 서비스: 바이오리포지토리 관리, 생물자원 기탁 서비스, 배양 등
- ISO 표준 서비스: 생물자원관리 표준, ATCC 표준 개발, 품질 관리, 생물자원 문서화
- 기술 및 교육 지원



〈그림 5〉 WDCM의 데이터베이스



〈그림 6〉 바이오포지터리 서비스

4.2.2 국내사례

국내의 경우 생명공학분야 뿐만 아니라 실험, 관측, 분석 등 연구개발 과정에서 산출된 연구 데이터는 다양한 활용 가치를 보유하고 있으나 연구성과물로 포함되고 있지 않아서 관리대상이 되지 않고 있는 것이 현실이다. 연구자 대부분이 연구데이터를 개인 자산으로 간주하고 개인적으로 관리하고 있으며 관리 공유에 대한 연구현장 인식이 저조한 상황이다(최명석, 이승복, 이상환 2017).

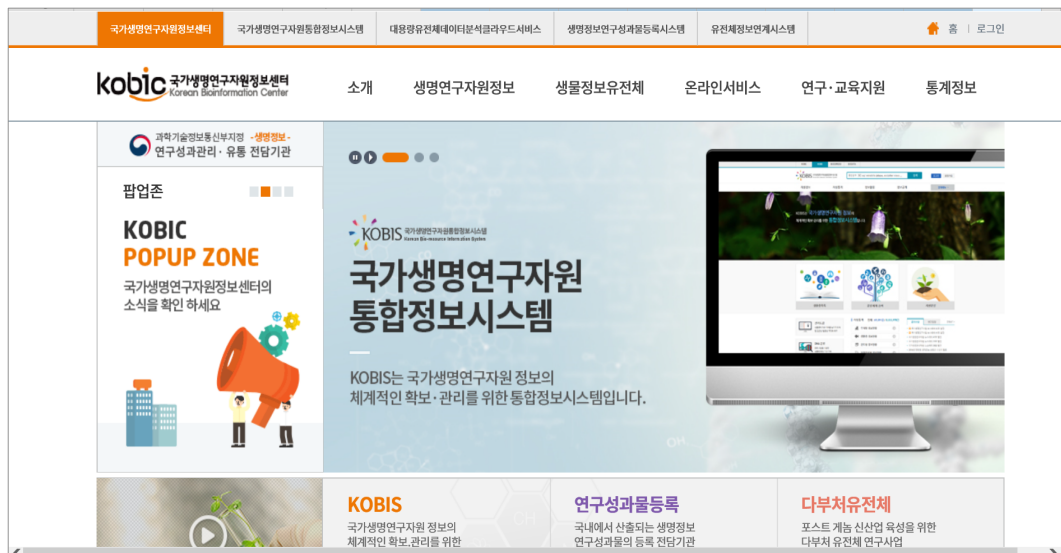
국내는 국외와 달리 분야별 공유 활용할 수 있는 전문 리포지터리 활용체계가 구축되어져 있지 않다. 단지 국내 생명공학분야의 대표 리포지터리인 국가센터인 국가생명연구자원정보센터(Korean Bioinformation Center; KOBIC)가 2010년 설립되어 부처별, 연구기관별로 흩어져 있는 생물자원, 생물다양성 그리고 생명정보를 범부처 차원에서 통합 관리하고 있는 정도이다. 이 센터에서는 연구데이터 관리측면에서는 대량의 데이터를 국제적으로 공인된 데이터 등록

기관(NCBI, EBI, DDBJ 등)에 제출 및 등록을 대행해 주는 생명정보 연구성과물 등록시스템 등이 있다. 국내사례로 이와 관련된 세부내용을 살펴보면 다음과 같다.

- 1) 국가생명연구자원정보센터(KOBIC: <https://www.kobic.re.kr/>)

국가생명연구자원정보센터는 부처별, 연구기관별로 흩어져 있는 생물자원, 생물다양성 그리고 생명정보를 통합 관리함으로써 국가적 차원에서 생물연구자원의 효율적인 활용이 이루어질 수 있도록, 정보연계 시스템을 구축하고 있다. 제공정보는 다음과 같다.

- 생명연구자원정보: KOBIS 소개와 연구자원의 대상, 범위, 구성에 대한 자원등록 연계기준을 제시하고 있다.
- 생물정보유전체: 유전체 정보를 비롯하여 다양한 종류의 대용량 오믹스 데이터¹⁰⁾를 체계적으로 수집하기 위한 인프라 구축하고, 다양한 분석 프로그램의 개발 및 서버



〈그림 7〉 KOBIC 홈페이지 서비스

스를 제공하고 있다.

- 온라인서비스: 데이터베이스 서비스로서 게놈, Genome, Proteome, Transcriptome, Structural Biology, Systems Biology 분야의 논문과 웹사이트를 소개하고 있다. 클라우드서비스로는 대용량 분석 서버나 분석 기술이 필요한 연구자들을 위하여 Bio-Express 대용량 유전체 데이터 분석 클라우드 서비스를 제공하고 있다.
- 연구교육지원: 주로 대용량 유전체 데이터, 단백질 구조 등의 bioinformatics 분석 지원과 데이터의 NCBI 등록 지원, 생명정보 프로그램 설치 자문 등의 서비스를 제공한다.

2) 생명정보 연구성과물 등록시스템
(<https://www.biodata.kr/>)

국가생명연구자원정보센터(KOBIC) 생명정보 연구성과물 등록시스템에서는 국내에서 산출되는 생명정보 연구성과물의 등록을 받는 전담기관으로써, 연구과제 정보는 국가과학기술 지식정보서비스(NTIS) 공통검색 API를 제공 받으며, 등록받은 연구성과물(생명정보)은 국가과학기술지식정보서비스(NTIS)와 연계되고 있다.

등록의뢰자의 요청에 따라서 KOBIC에서는 다음과 같은 서비스를 제공한다.

- 대량의 데이터(EST, GSS, Microarray, Whole Genome, RNA-Seq 등)를 국제적으로 공인된 데이터 등록기관(NCBI, EBI,

10) 인체자원의 유전체(genome), 단백질체(proteome) 등 생명정보를 분석하여 대량으로 생산·수집된 정보로서 보건의료 빅데이터는 2017년까지 개방형 플랫폼 구축을 목표로, 정보시스템간의 연계체계를 구축하고 개인정보 보호와 관리에 관한 법적근거와 관리체계를 마련할 계획이다.

- DDBJ 등)에 등록대행
 - 등록된 데이터에 대하여 KOBIC 자체 시스템을 활용한 분석 서비스 제공
 - 연구성과물(생명정보)의 국가과학기술지식정보서비스(NTIS)와 연계서비스
- 3) 생물자원센터(KTCT): 생물자원 연구 성과물기탁
(<http://biorp.kribb.re.kr/BiorpHome/>)

생물자원센터에서는 생물자원 기탁, 생물자원 분양 등의 서비스와 연구성과물 관리를 전담하고 있다. 국가연구개발사업으로 생산된 생물자원(미생물, 식물, 동물, 유전체(유전자클론, 플라스미드, 벡터 등)은 한국생명공학연구원 바이오의약 인프라사업부 생물자원센터(<http://biorp.kribb.re.kr>)에 기탁된다.

국내는 연구개발 사업을 통해 생산되거나 확보된 생물자원은 보존시설 및 기장비 부족, 전

문인력 부족 등으로 지속적인 보존관리가 어려웠으며 연구사업 종료에 따른 사후관리 체제의 미비로 활용이 활성화되지 못하고 있다.

수집된 연구성과의 검증과 분류, 가공 등을 통해 연구성과를 이용할 수 있도록 개방·공유하고 있으며 국가연구개발사업으로 생물자원(미생물, 식물, 동물, 유전체(유전자클론, 플라스미드, 벡터 등))을 발굴 및 국가연구 개발 성과물이 지속적으로 활용되도록 분야별 전담 리포지토리가 필요한 실정이다.

국내 사례를 살펴본 결과 생물공학분야 연구데이터 공유 활용을 위한 전문리포지토리는 아직 부재한 상태이며 국제적으로 공인된 데이터 등록기관(NCBI, EBI, DDBJ 등)에 등록대행 서비스를 제공하고 있는 정도이다. 앞으로 생명공학분야처럼 데이터 집약형 분야부터 전문 데이터센터를 구축하여 분야별 데이터 공유 활용이 이루어질 수 있도록 플랫폼 구축이 필요하다.



〈그림 8〉 연구성과물 관리

〈표 7〉 생명공학분야 연구데이터 공유사례 종합분석

	기관 리포지토리 기탁형		데이터뱅크 기탁형			전문데이터 센터 기탁형	데이터 아카이브 기탁형
	유럽생물 정보학 연구소	국가생명연구 자원정보센터	미국국립 생명공학 정보센터	생명정보 연구성과물 등록 시스템	생물 자원 센터	세계 미생물 데이터센터	미국 컬처 콜렉션
데이터관리계획 (DMP)	○	-	○	-	-	-	-
연구데이터 문서화	○	-	○	-	-	-	-
데이터 기술 (포맷)	○	-	○	-	-	-	-
데이터 저장	○	-	○	-	-	-	-
데이터 공유 및 접근 (사용)	○	-	○	-	-	-	-
데이터 보존	○	-	○	-	-	-	-

유럽생물정보학연구소와 미국 국립생명공학 정보센터에서는 연구데이터 공유 활용을 위해 연구데이터 관리계획, 문서화, 데이터포맷, 데이터 저장, 공유 및 접근, 보존 등 관리를 위한 각 단계를 시행하고 있는 반면, 국내 국가생명 연구자원정보센터, 중국 세계 미생물센터에서는 연구데이터 관리를 위한 DMP 외 각 단계가 시행되고 있지 않다. 단, 국내의 경우 연구성과물 중 생물자원에 대한 제출만 이루어지고 있다. 전문 리포지토리의 규모가 아닌 소규모의 데이터센터와 데이터 아카이브에서는 연구데이터 관리는 이루어지지 않고 있다.

5. 결론 및 제언

과학적 연구 과정은 연구데이터의 관리와 공유를 통해 향상된다. 올바른 데이터 관리는 신뢰할 수 있는 연구 결과의 검증과 기존 정보를 이용한 새롭고 혁신적인 연구를 가능하게 한다.

그것은 공적으로 연구를 지원하는 가치가 충분히 실현되기 위해서도 중요하다. 본 연구에서는 연구데이터 관리 사례로 미국 NISO DMP, 영국 UK 아카이브 DMP 등을 비교 분석하였고 1) 연구데이터 관리의 정책 계획 지원, 2) 연구데이터 관리 기술적 지원, 3) 연구데이터 공유 활용 지원, 4) 연구데이터 법적 메커니즘 지원, 5) 연구데이터 교육지원 등 5가지의 DMP 단계를 최종 도출하였다.

5가지 DMP 단계를 기준으로 DMP 운용 방식에 따라 생명공학분야의 연구데이터 공유방법으로 기관 리포지토리 기탁방법, 데이터 뱅크 기탁방법, 전문데이터 센터 기탁 방법으로 분류하였다. 연구데이터 분류방법에 따라 국내의 생명공학분야의 연구데이터 공유 활용 사례를 분석함으로써 국내 연구데이터 도입방안과 시사점을 도출하였다.

국내외 사례 분석결과 유럽생물정보학연구소와 미국 국립생명공학정보센터에서는 연구데이터 공유 활용을 위해 연구데이터 관리계획, 문서

화, 데이터포맷, 데이터 저장, 공유 및 접근, 보존 등 관리를 위한 각 단계를 시행하고 있는 반면, 국내 국가생명연구자원정보센터, 중국 세계 미생물센터에서는 연구데이터 관리를 위한 DMP 외 각 단계가 시행되고 있지 않다. 단, 국내의 경우 연구성과물 중 생물자원에 대한 제출만 이루어지고 있다. 전문 리포지토리의 규모가 아닌 소규모의 데이터센터와 데이터 아카이브에서는 연구데이터 관리는 이루어지지 않고 있다.

국내의 생명공학분야 데이터관리기관에서 연구데이터 공유 활용에 대한 서비스를 분석한 결과는 다음과 같다.

- ① 국외의 경우 오랫동안 분야별 전문리포지토리가 구축되고 있다. 즉, 생명연구자원 특성에 맞도록 공유 활용 플랫폼이 구축되어 있다. 그 형태는 분야별 전문 리포지토리, 데이터 뱅크 등 다양한 공유 활용 체계의 형태로 이루어지고 있다. 이 선도 사례를 통하여 연구데이터를 계획, 제출, 관리 등 공유 활용 생태계가 구축되어 있음을 확인할 수 있다.
- ② 국내의 경우 “국가연구개발사업의 관리 등에 관한 규정(대통령령 제 28210호 제 25조 13항, 2017.07.26.)” 및 “연구성과 관리·유통 전담기관 지정 고시(과학기술정보통신부 고시 제 2014-74호, 2014.11.03.)”에 따

라 한국생명공학연구원이 국가연구개발사업 9대 성과 중 생명자원(생물자원 및 생명정보) 연구성과 관리 유통 전담기관으로 지정되어있다.

따라서 국가연구개발사업으로 생산된 생물자원(미생물, 식물, 동물, 유전체(유전자클론, 플라스미드, 벡터 등))은 한국생명공학연구원 바이오의약인프라사업부 생물자원센터에 기탁된다. 수집된 생물자원에 대한 연구성과의 검증과 분류, 가공 등을 추진하고 있는 상황이지만 연구데이터는 연구성과물에 포함되어 있지 않고 관리되고 있지 않다. 따라서 연구데이터를 연구성과물에 추가하고 연구데이터 소유권을 규정하는 등 국가연구개발사업 관리 등에 관한 규정을 개정하는 등 국가차원의 연구데이터 관리 및 공유 활용체계 구축을 위한 제도적 기반 마련이 필요하다. 제도적 기반이 마련된 이후 연구데이터 관리계획(DMP)지원, 기술적 지원, 공유 활용 체계 지원, 법적 저작권 지원 등에 대한 가이드 매뉴얼 제작 배포가 필요하다.

본 논문에 국내외 연구데이터 공유 활용지원 선진사례를 보았는데 차후 연구에서는 연구데이터 정책 계획 지원, 연구데이터 기술적 지원, 연구데이터 법적 메커니즘 지원 등을 순차적으로 다루어볼 필요가 있다.

참 고 문 헌

- 강주연. 2017. 『생명공학분야 연구데이터 관리 방안 연구』. 석사학위논문. 전북대학교 대학원, 문헌정보학과.
- 김규빈. 2015. 과학데이터관리에 관한 해외 법제의 비교법적 고찰. 『과학기술과 법』, 6(1): 1-38.

- 김문정, 김성희. 2015. 과학기술분야 연구자의 연구데이터 공유의 영향요인에 대한 연구. 『한국문헌정보학회지』, 49(2): 313-334.
- 김선태. 2012. 『해양관측분야 메타데이터 표준요소 선정 및 설계에 관한 연구』. 박사학위논문. 전북대학교 대학원, 문헌정보학과.
- 김선태, 한선화, 이태영, 김용. 2010. 과학데이터 보존 및 활용모델에 관한 연구. 『한국비블리아학회지』, 21(4): 81-93.
- 김은정, 남태우. 2012. 연구데이터 수집에 영향을 미치는 요인 분석. 『정보관리학회지』, 29(2): 27-44.
- 김지현. 2011. 자연과학분야 대학실험실에서의 연구노트 작성 및 관리에 관한 연구. 『한국기록관리학회지』, 11(1): 139-159.
- 김지현. 2012. 대학 내 연구자들의 연구데이터 관리에 관한 연구. 『한국도서관·정보학회지』, 43(3): 433-455.
- 김지현. 2013. 국외 정부연구비지원기관의 연구데이터 관리정책 분석: 미국, 영국, 캐나다, 호주를 중심으로. 『한국문헌정보학회지』, 47(3): 251-274.
- 김지현. 2014. 대학도서관의 연구데이터관리서비스에 관한 연구: 미국 연구중심대학도서관을 중심으로. 『한국비블리아학회지』, 25(3): 165-189.
- 김지현. 2015. 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 『한국문헌정보학회지』, 47(3): 413-436.
- 김지현. 2016. 연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구. 『한국도서관·정보학회지』, 47(3): 213-239.
- 윤종민, 김규빈. 2013. 과학데이터에 관한 입법례와 관리정책 그리고 대응방안: 호주, 미국, 중국을 중심으로. 『한국기술혁신학회지』, 16(1): 63-100.
- 최명석, 이승복, 이상환. 2017. 국내 과학기술분야 연구기관의 과학데이터 관리 현황. 『한국콘텐츠학회논문지』, 17(12): 117-126.
- Eschenfelder, K. R. and A. Johnson. 2014. "Managing the data commons: Controlled sharing of scholarly data." *Journal of the Association for Information Science and Technology*, 65(9): 1757-1774.
- Eynden, Veerle Van den, Matthew Louise Corti, Libby Bishop Woollard, and Laurence Horton. 2011. *Managing and Sharing Data: Best Practice For Researchers*, Third Edition. UK: UK Data Archive.
- OECD. 2007. *OECD Best Practice Guideline for Biological Resource Centres*. France: OECD.
- UK Data Archive. 2011. *Managing and Sharing Data: Best Practice For Researchers*. UK: UK Data Archive.

• 국문 참고자료의 영어 표기

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

- Choi, Myung-Seok, Seung-Bock Lee, and Sanghwan Lee. 2017. "Research Data Management of Science and Technology Research Institutes in Korea." *Journal of the Korea Contents Association*, 17(12): 117-126.
- Gang, Ju-yeon. 2017. *A Study on Analysis of the Current Status and the Method to Improve Biotechnology Research Data Management*. Master thesis, Dept. of Library and information Science, The Graduate School Chonbuk National University.
- Kim, Eun-Jung and Tae-Woo Nam. 2012. "Factor Analysis of Effects on Research Data Collection." *Korean Society for Information Management*, 29(2): 27-44.
- Kim, Jihyun. 2011. "A Study on the Creation and Maintenance of Laboratory Notebooks in Scientific Laboratories of a University." *Korean Society of Archives and Records Management*, 11(1): 139-159.
- Kim, Jihyun. 2012. "A Study on University Researchers' Data Management Practices." *Korean Library and Information Science Society*, 43(3): 433-455.
- Kim, Jihyun. 2013. "An Analysis of Data Management Policies of Governmental Funding Agencies in the U.S., the U.K., Canada and Australia." *Korean Society for Library and Information Science*, 47(3): 251-274.
- Kim, Jihyun. 2014. "A Study on Research Data Management Services of Research University Libraries in the U.S." *Korean Bibliography Society for Library and Information Science*, 25(3): 165-189.
- Kim, Jihyun. 2015. "A Study on the Perceptions of University Researchers on Data Management and Sharing." *Korean Society for Library and Information Science*, 47(3): 413-436.
- Kim, Jihyun. 2016. "A Study on Policy Components of Data Access and Use Controls in Research Data Repositories." *Korean Library and Information Science Society*, 47(3): 213-239.
- Kim, kyu bin. 2015. "A Comparative Legal Analysis of Research Data Management." *Science, Technology and Law*, 6(1): 1-38.
- Kim, Moonjeong and Seonghee Kim. 2015. "A Study on the Factors Affecting Sharing of Research Data of Science and Technology Researchers." *Korean Society for Library and Information Science*, 49(2): 313-334.
- Kim, Sun-Tae. 2012. *A Study on Extraction and Design of Standardized Elements on Metadata*

for Ocean Observational Dat. Ph. D. diss., Dept. of Library and information Science, The Graduate School Chonbuk National University.

Kim, Sun-Tae, Sun-Hwa Han, Tae-Young Lee, and Yong Kim. 2010. "A Study on a Model for Using and Preserving Scientific Data." *Korean Biblia Society for Library and Information Science*, 21(4): 81-93.

Yoon, Chong-min and Kyubin Kim. 2013. "Legislation Cases, Management Policies and Countermeasures on Scientific Data - Focusing Australia, the United States and China." *Korea Technology Innovation Society*, 16(1): 63-100.