

연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구

A Study on Policy Components of Data Access and Use Controls in Research Data Repositories

김 지 현(Jihyun Kim)*

〈목 차〉

I. 서 론	1. 조사대상 레포지터리의 특성
II. 연구데이터 공유와 접근	2. 정책 내용 분석
1. 연구데이터 공유 방식	3. 분석 결과 요약
2. 연구데이터 접근 및 이용 정책	IV. 결론 및 제언
III. 데이터 레포지터리의 접근 및 이용정책 분석	

초 록

전 세계적으로 오픈 데이터가 강조되는 환경에서 데이터의 공유 및 재이용으로 인해 발생하는 문제점을 최소화하기 위한 정책적 논의도 함께 진행되고 있다. 본 연구에서는 연구데이터 레포지터리에서 데이터의 접근 및 이용을 통제하는 정책요소들을 조사하고 학문 분야별로 그러한 정책 요소들의 공통점과 차이점을 살펴보는 것을 목적으로 하였다. 이를 위해 해외 연구데이터 레포지터리 37곳을 대상으로 데이터 접근 및 이용 통제를 규정하고 있는 정책 요소를 분석하였다. 생명과학 및 보건과학 분야 20개 레포지터리, 화학·지구환경과학·물리학 분야 10개 레포지터리, 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 나타났으며 이는 분야별로 강조되는 데이터 접근 및 이용 통제 근거의 차이를 반영하는 것으로 볼 수 있다.

키워드: 연구데이터 레포지터리, 접근 통제, 이용 통제, 데이터 정책

ABSTRACT

As Open Data has been emphasized globally, discussions on data policies have occurred for minimizing problems resulting from data sharing and reuse. This study aimed at investigating policy components for controlling access and use of data and examining similarities and differences of the policy components across disciplines. For the purpose, the study analyzed policy components for data access and use controls provided by 37 research data repositories overseas. These included twenty repositories in biological and health science, ten in chemistry, earth and environmental science and physics, as well as seven in social science and general science. The analysis showed that common policy components involve copyright/licenses, data citation, disclaimers and embargoes. However, there were differences in diversity of policy components among the disciplines and it indicated that the rationales of access and use controls emphasized would be different in the disciplines.

Keywords: Research data repositories, Access control, Use control, Data policies

* 이화여자대학교 사회과학대학 문헌정보학전공 조교수(kim.jh@ewha.ac.kr)

•논문접수: 2016년 8월 20일 •최초심사: 2016년 8월 30일 •게재확정: 2016년 9월 6일

•한국도서관·정보학회지 47(3), 213-239, 2016. [<http://dx.doi.org/10.16981/kliss.47.201609.213>]

I. 서론

지식의 확산과 혁신적인 연구개발에 있어 오픈 액세스(Open Access, 이하 OA)는 전 세계적으로 수용되고 있는 핵심 전략 중 하나이다. 특히 공공기금으로 수행된 연구의 결과물은 공공재(public goods)의 성격을 지니므로 누구나 무료로 접근 제한 없이 이러한 결과물을 활용하게 하려는 노력이 다양한 차원에서 전개되고 있다. 이와 같은 활동의 일환으로 최근 OECD(Organization for Economic Co-operation Development)에서는 공공연구의 결과물인 학술지 논문과 연구데이터의 OA를 강조하는 오픈 사이언스 프로젝트(Open Science Project)를 수행하였다. 오픈 사이언스는 “과학계에서 연구 자료를 공개·검증하는 개방적 연구 규범을 지칭”하며 “개방적 연구협력을 도모하는 다양한 활동과 과학의 긍정적 외부효과를 제고하기 위한 활동”이다(신은정 2015, 2). 오픈 사이언스 프로젝트는 연구 결과물로의 접근성 향상을 통해 오픈 사이언스 활동을 지원하는 하나의 시도라고 할 수 있다.

한국을 비롯한 OECD 회원국들의 오픈 사이언스 정책 동향 및 활동을 조사한 OECD의 2015년 보고서에 따르면 공적자금으로 수행된 연구 출판물의 OA와 그러한 연구에서 생성된 데이터를 공개하는 오픈 데이터(Open Data)가 오픈 사이언스의 근간임을 제시하고 있다. 이 중에서 오픈 데이터는 누구든지 기술적 혹은 법적 제한 없이 공적자금으로 수행된 연구에서 생성된 데이터에 접근하고 재이용(reuse)할 수 있게 하는 것을 말한다. OECD의 국가별 현황 조사에 의하면 출판물의 OA 정책은 보편적으로 수립되어 있는 반면 오픈 데이터 정책 수립에는 어려움이 많은 것으로 나타났다. 이러한 어려움에는 데이터 소유권 파악의 어려움, 개인정보보호, 비밀유지 및 보안 문제, 데이터의 품질 차이, 연구자 커뮤니티에서 데이터 관리 능력과 오픈 데이터에 대한 인센티브 부족 등이 있다(OECD 2015).

한 예로 보건·생명과학 분야의 대표적인 공공연구기금 지원기관인 미국 국립보건원(National Institute of Health, NIH)의 경우 2000년대 초반부터 데이터의 공유를 강조해왔다. 그러나 NIH에서 지원하는 연구과제로부터 생성된 데이터 중 유전체 데이터(genomic data)와 같이 민감한 데이터의 경우 무조건적인 OA는 사실상 지양되어야 하며 데이터 공유와 관련된 여러 문제점들을 최소화할 수 있는 데이터 접근 및 이용 정책이 필요하다는 목소리가 높다(Gitter 2011; Paltoo et al. 2013; Shabani et al. 2015). 또한 연구데이터 레포지터리 입장에서는 데이터 접근 및 이용을 통제함으로써 데이터의 무결성(integrity)을 강화할 수 있고 이를 통해 레포지터리의 지속가능성과 데이터의 기탁 비율을 높일 수 있다는 주장도 제기되고 있다(Eschenfelder와 Johnson 2014).

본 연구는 오픈 데이터가 강조되는 현실에서 연구데이터 레포지터리에서는 데이터의 OA

로 인해 발생할 수 있는 문제점들을 어떻게 정책적으로 통제하고 있는지를 살펴보고, 학문 분야별로 공통점과 차이점을 논의하는 것을 목적으로 하였다. 이를 위해 학술지 Nature에서 발행하는 데이터 저널(data journal)인 “Scientific Data”에서 추천하고 있는 연구데이터 레포지터리 82개 중 데이터 접근 및 이용 정책 또는 관련 내용을 제공하는 37개 레포지터리를 선정하였다. 이들 37개 레포지터리에서 제공하는 정책 내용을 기초로 데이터의 접근 또는 이용을 통제하는 정책 요소들을 분석하였다. 학문분야별로는 해당 연구데이터 레포지터리를 생명과학 및 보건과학, 화학, 지구환경과학 및 물리학, 사회과학 및 일반과학의 세 가지 영역으로 구분하여 정책 요소들의 특성을 살펴보았다. 데이터 공유와 재이용으로 인한 우려를 완화할 수 있는 정책적인 요소를 살펴봄으로써 오픈 데이터와 지속가능한 데이터 활용의 조화를 이루는 정책적 논의의 필요성을 제안하는 것에 본 연구의 의의가 있다.

II. 연구데이터 공유와 접근

1. 연구데이터 공유 방식

연구데이터의 공유와 재이용에 대한 요구가 높아지면서 데이터의 접근과 이를 지속가능하게 하는 관리적 요건들을 포함하는 연구데이터 공유 방식에 대한 논의가 이루어져왔다. 현재까지 실행되고 있는 데이터의 공유, 보존 및 접근을 지원하는 방식은 크게 두 가지로 연구데이터 레포지터리(Research Data Repository, RDR)와 데이터 저널(data journal)이 있다. 연구데이터 레포지터리는 데이터를 장기적이고 신뢰할 수 있는 방식으로 저장하고 접근할 수 있게 하는 정보 인프라를 의미한다. 여기에는 다양한 형태의 데이터 아카이브와 데이터 센터 또는 디지털 도서관이 포함된다(Pompel et al. 2013).

데이터 저널(data journal)은 데이터를 생산하는 연구자들이 공식적으로 데이터를 출판하여 공개하고 타 연구자들이 그 데이터를 인용함으로써 연구결과에 대한 인정(acknowledgement)을 받을 수 있는 통로로 활용하기 위해 개발된 것이다(ANDS 2016). 데이터 저널은 기존의 학술지 출판 모형을 데이터 공유에 접목시킨 것으로 보상체계의 부족과 데이터 인용 및 품질 관리 문제를 완화할 수 있는 방식으로 그 수가 증가하고 있는 추세이다(Candela et al. 2015). 그러나 데이터 저널을 통해 출판되는 내용은 데이터에 대한 기술(description)이며 데이터 자체에 대한 접근과 활용을 위해서는 궁극적으로 데이터를 장기적으로 접근 가능하게 하는 신뢰성 있는 레포지터리를 필요로 한다. 이러한 데이터 레포지터리는 다시 분야별 레포지터리(disciplinary repository), 기관 레포지터리(institutional repository), 그리고 연구과제 중심

형(project-specific) 레포지터리로 구분할 수 있다(Pompel과 Dalmeire-Tiessen 2013). 따라서 본 연구에서는 연구데이터 레포지터리를 데이터 공유의 기본적인 인프라로 간주하고 데이터 레포지터리에서 제공하는 접근 및 이용 정책을 분석하였다.

2. 연구데이터 접근 및 이용 정책

오픈 데이터 정책 및 전략 수립 동향을 논의한 Nicol et al.(2013)은 유럽연구지대(European Research Area, ERA)에 속한 국가들과 브라질, 캐나다, 일본, 미국의 2000여 개 정부기관, 연구비지원기관, 연구기관 및 대학을 대상으로 설문조사를 실시하였다. 정부기관에서는 오픈 데이터가 정보기반 경제사회에서 경쟁 우위를 제공하여 새로운 상품과 서비스 개발의 초석이 될 것이라고 전망하고 있었다. 데이터를 통해 수익을 창출하는 기관들도 있었다. 그러나 데이터와 메타데이터 포맷 표준화의 어려움과 디렉토리 및 색인의 부족으로 인한 검색과 접근의 어려움이 문제점으로 제시되었다.

Eschenfelder와 Johnson(2014)는 연구데이터의 접근 및 이용 통제 정책(access and use control policies)을 제공하는 24개 데이터 레포지터리를 대상으로 정책의 내용 분석과 레포지터리 관계자와의 설문 및 인터뷰를 수행하였다. 이 연구에서는 OA를 표방하는 데이터 레포지터리는 많지만 그에 못지않게 데이터의 접근(누가 데이터를 이용하는가)과 이용(어떻게 그들이 데이터를 재이용하는가)에 통제를 가하는 다수의 데이터 레포지터리가 존재한다는 사실에 주목하였다. 접근 및 이용 통제는 레포지터리 운영의 범위를 정하고 정책수립 및 데이터 재이용의 법적·윤리적 측면을 다루는 데 있어 중요성을 가지므로 이를 분석할 필요가 있다고 보았다. 화학, 지구과학, 생태학, 사회과학, 인문학 분야 레포지터리의 정책 및 운영현황 분석을 통해 1) 접근 및 이용 통제 정책의 수립과 승인 주체; 2) 레포지터리의 사명; 3) 민감한 데이터의 관리; 4) 데이터 이용의 의미; 5) 레포지터리 내에서의 데이터 접근 및 이용 통제 방식의 다양성 정도를 데이터 접근과 이용 통제의 다섯 가지 측면으로 제시하였다.

Doshi et al.(2016)은 미국 내 연방정부 및 주정부에서 운영되면서 공적자금의 지원을 받는 보건 의료 분야 데이터베이스에서 연구 목적으로 널리 활용되는 데이터세트 116개를 대상으로 이에 대한 접근 정책을 분석하였다. 정책의 내용은 연방정부기관 간 그리고 주정부 간에 매우 다양한 양상을 보였으며 하나의 기관 데이터베이스에서 제공되는 여러 개의 데이터세트 간에도 접근 정책의 내용이 상이한 것으로 나타났다. 데이터 활용 신청 절차의 복잡성, 높은 이용료 및 신청한 데이터를 받기까지 상당한 시간이 걸리는 등 데이터 활용을 방해하는 요인들도 지적되었다.

보건 의료 분야의 데이터 중 임상 시험(clinical trail) 데이터는 OA로 공개함에 있어 여러 가지 제약이 따른다. Sydes et al.(2015)는 임상 시험 데이터 공유의 어려움으로 1) 임상

시험 동의서에 데이터를 인터넷 상에 공개한다는 내용을 명시하지 않는 경우가 많다는 점; 2) 임상 시험 참여자의 신원이 노출될 수 있다는 점; 3) 데이터 익명화로 인한 데이터 왜곡 발생 우려; 4) 데이터가 부적절한 목적으로 남용될 수 있다는 점을 들었다. 이러한 어려움으로 인해 임상 시험 데이터는 통제된 접근을 허용하는 것이 바람직하다는 의견을 제시하였으며 통제된 접근의 근거를 다음과 같이 제안하였다. 첫째, 연구자가 데이터 재이용을 요청할 때는 합당한 근거가 있어야 한다. 둘째, 원래의 임상 시험에 해가 된다면 데이터를 공유하지 않는다. 셋째, 원래의 임상 시험 연구자들은 데이터를 독점적으로 사용할 수 있는 기간을 부여받아야 한다. 넷째, 데이터를 공유하는데 필요한 적절한 수준의 자원이 제공되어야 한다. 다섯째, 모든 데이터 공유는 각 국가의 데이터 보호 정책을 준수하는 범위 내에서 이루어져야 한다. 여섯째, 모든 임상 시험은 명성 있는 임상 시험 데이터베이스에 등록되어 검색 가능해야 한다.

연구데이터의 통제된 접근에 대한 논의는 유전체학(genomics) 분야에서도 활발히 이루어져왔다. Gitter(2011)는 공적자금으로 수행된 연구에서 생산된 데이터의 OA 정책을 수립하고 실행한 사례로서 1000 유전체 프로젝트(1000 Genome Project)의 데이터 접근 정책을 분석하고 문제점을 제시하였다. 해당 정책에서는 프로젝트에서 생산된 모든 유전형(genotype) 데이터를 공개하되 데이터 생산자들은 학술발표와 학술지 논문 출판에 있어 우선권을 가진다는 것을 명시하였다. 그러나 데이터 생산자들은 데이터 공유를 통해 발생할 수 있는 연구 아이디어의 도용과 데이터의 오용에 대한 우려로 인해 다른 연구자들의 데이터 접근을 허용하는 것을 주저하는 경향이 많았다. 이러한 우려를 최소화하는 방안으로 데이터를 독점적으로 활용할 수 있는 엠바고 기간을 제공할 필요가 있음을 제시하였으나 데이터 공유를 촉진시킬 수 있는 적절한 기간을 정의하는 일 역시 쉽지 않은 것으로 나타났다. 또한 국제적인 범위의 프로젝트의 경우 각 국가에서 다양한 연구지원기관들이 참여하게 되는데 데이터 공유와 보존을 위한 인프라 구축에는 아무도 책임지려하지 않는 문제점도 제기되었다. 이와 같은 사례 분석을 통해 소규모 과학(small science) 연구에서는 오픈 데이터보다 공동연구자들 간의 제한적인 데이터 공유가 유용한 방안임을 제시하였다.

Ramos et al.(2013)은 유전체학 데이터 공유 사례로서 미국 국립생물공학정보센터(National Center for Biotechnology Information, NCBI)에서 개발한 dbGaP(the database of Genotypes and Phenotypes)사례를 분석하였다. dbGaP를 통해 GAIN(Genetic Association Information Network)의 지원을 받아 수행된 전장 유전체 연관 분석(Genome-wide association studies, GWAS) 데이터에 접근하는 방식을 논의하였는데 접근의 수준은 크게 두 가지로 OA와 통제된 접근으로 나뉜다. GWAS 데이터가 민감한 내용을 포함하고 있지 않은 경우 OA가 가능하지만 그렇지 않은 경우 GAIN 데이터 접근 위원회(Data Access Committee, DAC)로부터 접근에 대한 승인을 받아야 한다.

Paltoo et al.(2013) 역시 dbGAP를 통한 GWAS 데이터 공유에 대해 논의하였는데 실험 참가자들의 동의서에 기초를 둔 통제된 접근 방식을 제안하였다. 유전체 데이터의 윤리적인 활용과 일반 대중의 윤리적인 데이터 활용에 대한 인식은 여전히 중요한 과제이며 향후 미국 국립보건원(National Institute of Health, NIH)을 중심으로 하는 접근 정책 수립이 활발히 이루어질 것이라 전망하였다.

Kaye와 Hawkins(2014)는 유전체학 분야에서의 데이터 공유 관리체계를 논의하고 데이터 접근 정책 개발의 문제점을 논의하였다. 그러한 문제점으로 1) 데이터 접근을 데이터의 유형이 아닌 연구과제 단위로 관리한다는 점; 2) 연구 참여자들의 데이터 공유에 대한 동의가 명시되지 않은 경우가 많다는 점; 3) 데이터 공유에 대한 인정 및 보상 체계의 부족; 4) 데이터 공유에 소요되는 관리적인 비용의 문제를 제시하였다. 지속가능한 데이터 공유체계 마련을 위해서는 데이터, 정책 및 절차, 데이터를 이용하는 연구자, 데이터 접근을 지원하는 인프라 등 다양한 차원에서의 논의가 필요하다고 보았다.

Shabani et al.(2015)는 유전체학 분야에서 접근 및 이용 통제 정책 수립이 증가하는 현상에 주목하면서 접근 정책 수립의 핵심은 데이터 접근 방법의 구체적인 명시와 데이터 접근 위원회 운영이라고 언급하였다. 데이터 이용자와 연구비지원기관의 의견을 반영하고 연구 참여자들의 개인정보 유출 우려를 최소화하는 데이터 접근 정책 수립이 필요하며 데이터 접근 위원회와 연구 윤리 위원회 간의 관계와 각 위원회의 역할에 대한 정립이 필요함을 제안하였다.

선행연구를 통해 임상 시험 데이터 또는 유전체 데이터와 같이 민감한 데이터의 접근 및 이용 제한을 정책으로 수립하고 데이터 접근 위원회와 같은 통제 기제의 적절한 운영이 논의되고 있음을 알 수 있다. 비록 보건 생명과학 분야의 특정 데이터에 한정된 논의이기는 하지만 인간을 대상으로 하는 사회과학 연구에서 생산되는 데이터나 비밀유지 혹은 상업적 이해관계가 중요한 데이터 등 타 분야의 데이터 접근과 활용에도 참고사항이 될 수 있다. 다음 장에서는 해외 연구데이터 레포지터리 정책에서 제시된 데이터 접근 및 이용 통제 요소들을 살펴보고자 한다.

Ⅲ. 데이터 레포지터리의 접근 및 이용 정책 분석

1. 조사 대상 레포지터리의 특성

본 연구에서는 연구데이터 레포지터리의 접근 및 이용정책을 분석하기 위해 우선 학문 분야별로 모범사례로 알려진 데이터 레포지터리를 파악하여 접근 정책의 유무를 확인하였다. 모범사례로 인정되는 레포지터리의 경우 국제적으로 권위 있는 학술지인 Nature에서 출판하

는 데이터 저널인 Scientific Data에서 추천하는 레포지터리 82개를 대상으로 하였다.

Scientific Data는 온라인으로만 출판되고 동료심사(peer-review) 제도에 기반을 두고 있는 OA 데이터 저널로서 출판되는 내용은 과학적으로 가치 있는 데이터세트에 대한 기술(descriptions)이며 이를 데이터 디스크립터(data descriptors)라고 명명하고 있다. 데이터 디스크립터는 모두 14개의 구성요소로 이루어져 있는데 제목, 저자 및 소속기관, 초록, 배경 정보 및 요약, 데이터 생성 및 수집 방법론(methods), 데이터가 저장된 레포지터리 및 데이터 파일과 포맷에 대한 개요를 나타내는 데이터 기록(data records) 등을 포함한다. Scientific Data에 출판하려면 OA 학술지에서 주로 이루어지는 비용 모형인 논문처리비용(APC: Article-Processing Charges)을 지불해야 한다(Scientific Data 2016b).

Scientific Data에는 데이터에 대한 기술 정보가 출판되지만 실제 데이터의 저장과 관리는 데이터 레포지터리에서 이루어진다. 이를 위해 Scientific Data에서는 추천 데이터 레포지터리(recommended data repositories)¹⁾ 목록을 제공하여 저자들이 데이터 디스크립터를 제출하기 전에 추천 레포지터리 목록에서 원하는 레포지터리를 선택하여 데이터를 기탁하게 하고 있다. Scientific Data에서 제공하는 추천 레포지터리의 선정 기준은 다음의 7가지인데 1) 학문 분야별 커뮤니티에서 널리 알려진 것인지; 2) 데이터 세트의 장기적인 보존이 가능한지; 3) 전문적인 큐레이션 서비스를 제공하는지; 4) 커뮤니티에서 인정되고 적절한 형태의 필수적인 보고 항목(reporting requirements)을 제공하고 있는지; 5) 데이터의 공개에 앞서 익명의 동료 평가자들이 데이터에 접근할 수 있게 허용하는지; 6) 기탁된 데이터 세트에 식별 기호(identifier)를 제공하는지; 7) 불필요한 제한 없이 데이터에 대한 공적인 접근(public access)을 허용하는지를 포함한다. 또한 저자들로부터 데이터 레포지터리를 추천받아 적합하다고 판단될 경우 목록에 추가하기도 한다. 매년 분기별로 추천 레포지터리 목록을 갱신하고 있으며 최신 목록은 2016년 5월에 제공된 것이다(Scientific Data 2016a).

본 연구에서는 Scientific Data의 최신 데이터 레포지터리 추천 목록에 포함된 82개의 레포지터리를 살펴보고 이 중에서 데이터 접근 및 이용 정책 혹은 접근이나 이용과 관련된 내용을 포함하는 메뉴를 제공하는 레포지터리를 확인하였다. 그 결과 총 37개(45.1%)의 레포지터리에서 접근 및 이용 정책 또는 관련 내용을 포함하는 메뉴를 제공하고 있었다. <표 1>에서와 같이 추천 레포지터리 82개는 학문 분야별로 분류되어 있는데 생명과학의 경우 핵산 서열, 단백질 서열 등 9개의 세부 영역으로 나누어져 있다. 조사 대상으로 선정된 37개 레포지터리의 학문 분야별 분포를 보면 생명과학 분야의 단백질 서열, 분류와 종 다양성, 세포계측학 영역을 제외하고는 제시된 분야마다 적어도 1개 이상의 레포지터리에서 접근 및 이용 정책 또는 관련 내용을 제공하고 있음을 알 수 있다.

1) <http://www.nature.com/sdata/policies/repositories>

〈표 1〉 학문분야별 데이터 레포지터리 분포

학문 분야별 분류		추천 레포지터리 전체(n=82)	조사 대상 레포지 터리(n=37)
생명과학 (biological sciences)	핵산 서열(nucleic acid sequence)	10	1
	단백질 서열(protein sequence)	1	0
	분자 및 초분자 구조(molecular & supramolecular structure)	6	2
	신경과학(neuroscience)	3	2
	오믹스(omics)	11	5
	분류 & 종 다양성(taxonomy & species diversity)	4	0
	수학적 모델링(mathematical & modelling resources)	2	2
	세포계측학(cytometry)	1	0
	생명과학 커뮤니티(life-science community resources)	9	4
보건과학(Health sciences)		8	4
화학 및 화학생명과학(Chemistry & chemical biology)		3	1
지구환경과학(Earth and environmental sciences)		13	8
물리, 천체물리 및 천문학(Physics, astrophysics & astronomy)		4	1
사회과학(Social sciences)		5	5
일반과학(General science)		2	2

조사 대상인 37개 레포지터리는 <표 2>에 제시되어 있으며 레포지터리명이 길 경우 추천 목록에서 제시된 약어를 표시하였다. 레포지터리 웹사이트에서 제공되는 접근 및 이용정책 혹은 관련 내용이 제공되는 메뉴명을 살펴보면 6개 레포지터리에서 ‘약관 및 조건(terms and condition)’이라는 명칭을 활용하고 있었으며 또 다른 7개 레포지터리에서 ‘이용 약관(terms of use)’라는 명칭을 포함하여 사용하고 있었다. Eschenfelder와 Johnson(2014)의 연구에서도 주로 이러한 명칭이 데이터 레포지터리에서 접근 및 이용 정책을 지칭하는 것으로 활용되고 있음을 제시하였고 본 연구에서도 같은 현상을 확인할 수 있었다.

정책이라는 용어를 명시하고 있는 메뉴명 중에서 3개 레포지터리에서 ‘데이터 접근 정책(data access policy)’, 또 다른 3개 기관에서 ‘데이터 정책(data policy 또는 policies)’이라는 명칭을 사용하고 있었다. 그 외에도 데이터 이용가능성(data availability) 또는 통제 및 배포(hold and release) 정책이라는 명칭을 사용하는 경우도 있었으며 해당 레포지터리 정책 내에 데이터 접근 및 이용에 대한 내용을 포함하는 경우도 있었다. 데이터 제출 및 배포(Data submission and release) 혹은 이용이 제한된 데이터로의 접근(Access to restricted-use data)과 같은 명칭도 존재하였다.

약관 또는 정책이라는 용어가 포함되지 않은 메뉴에서도 접근 또는 이용과 관련된 내용을 제공하고 있었는데 2개 레포지터리에서 FAQ를 그러한 통로로 활용하고 있었으며 2개 레포지터리에서는 About 메뉴를 활용하고 있었다. 그 외에 행동 강령(Code of conduct), 협력적 협약(Collaborative agreements), 법적 고지(Legal notice)라는 메뉴에서도 데이터 접근 및 이용 관련 내용을 제시하고 있었다.

〈표 2〉 조사 대상 레포지터리 및 접근 및 이용정책 메뉴명

학문 분야별 분류		레포지터리*	레포지터리명 약어	접근 및 이용정책 메뉴명*
생명과학	핵산서열정보	EMBL Nucleotide Sequence Database	ENA	Data availability policy
	분자 및 초분자 구조	Protein Circular Dichroism Data Bank	PCDDDB	Terms and conditions
		Electron Microscopy Data Bank	EMDB	Hold and release policy
	신경과학	NeuroMorpho.org	-	Terms of use
		Functional Connectomes Project International Neuroimaging Data-Sharing Initiatives	FCP/INDI	Data access policy
	오믹스	ArrayExpress	-	Data access policy
		dbGAP	-	Code of conduct
		The European Genome-phenome Archive	EGA	FAQs
		Database of Interacting Proteins	DIP	Terms of use
		ProteomeXchange	-	Collaborative agreement
	수학적 모델링	BioModels Database	-	Terms of use
		Kinetic Models of Biological Systems	KiMoSys	Terms and condition of use
	생명과학 커뮤니티	Eukaryotic Pathogen Database Resources	EuPathDB	Data submission and release
		VectorBase	-	Data policies
		Xenbase	-	About Xenbase
		Biological General Repository for Interaction Datasets	BioGRID	Terms and condition
보건과학		National Addiction & HIV Data Archive Program	NAHDAP	Access to restricted-use data
		National Database for Autism Research	NDAR	NDAR policy
		Virtual Skeleton Database (SICAS medical image repository)	VSD	Terms and conditions
		National Database for Clinical Trials related to Mental Illness	NDCT	Policy for the NDA
화학 및 화학생명과학		ChEMBL	-	EBI terms of use
지구환경과학		Cold and Arid Regions Science Data Center (Lanzhou, China)	CARD	Conditions of use
		NERC Data Centres	NERC	NERC data policy
		LTER Network Information System Data Portal	LTER	Data access policy
		PANGAEA	-	About / imprint
		Australian Antarctic Data Centre	AADC	Conditions of use
		Marine Geosciences Data System	MGDS	Terms of use
		Seano	-	Legal notice
		NASAGoddardEarthSciencesDataandInformationServicesCenter	GES DISC	Data policy
물리, 천체물리 및 천문학		NoMaD Repository	NoMaD	Terms of use
사회과학		Harvard Dataverse	-	Harvard Dataverse policies
		openICSPR	-	FAQs
		UK Data Service - ReShare	-	Terms and condition of access
		Open Science Framework	OSF	Terms of use
		Archaeology Data Service	ADS	Collections policy
일반과학		Dryad Digital Repository	-	Terms of services
		figshare	-	Terms and conditions

*각 레포지터리와 접근 및 이용정책 웹페이지 URL은 <부록 1>에 제시하였음.

2. 정책 내용 분석

조사 대상 레포지터리에서 제공하는 접근 및 이용 정책과 관련 내용 분석을 위해 Eschenfelder와 Johnson(2014)에서 제시된 분석 기준을 활용하였다. 해당 선행 연구에서는 사회과학, 인문학, 보건과학, 생태학, 화학/분자생물학, 지구/우주과학 분야 24개 레포지터리의 접근 및 이용 정책을 분석하였는데 이들 24개 레포지터리와 본 연구의 조사대상 레포지터리 간 중복은 없는 것으로 나타났다. <표 3>의 왼쪽 열에 제시된 대로 선행 연구에서의 분석 기준은 이용 통제(use control)와 접근 통제(access control)의 두 영역에서 각각 11개와 7개가 제시되었는데 본 연구에서는 이를 중심으로 내용 분석을 시작하였다. 분석을 진

<표 3> 내용 분석 기준

선행연구에서 제시된 분석 기준	최종 분석 기준
이용 통제	
승인되지 않은 개인에 의한 데이터 배포 금지	승인되지 않은 개인에 의한 데이터 배포 금지
학술적 목적으로 제한 [†]	비상업적 목적으로 제한 ^{**}
인용 등 인정체계(acknowledgement)요청	인용 등 인정체계(acknowledgement)요청
데이터 이용에 관해 레포지터리나 데이터 생산자에게 보고 요청	데이터 이용에 관해 레포지터리나 데이터 생산자에게 보고 요청
승인된 연구가 아니면 데이터 배포 금지	승인된 연구가 아니면 데이터 배포 금지
데이터의 이용 후 혹은 특정 시기가 지나면 데이터 폐기	데이터의 이용 후 혹은 특정 시기가 지나면 데이터 폐기
데이터에 나타난 개인의 비밀 유지	데이터에 나타난 개인의 비밀 유지
데이터의 안전한 저장 요청 [†]	데이터 관련 보안 문제 발생 시 보고 요청
잘못된 결과를 초래할 수 있는 데이터 이용 금지 [†]	면책조항(disclaimer)*
데이터 관련 보안 문제 발생 시 보고 요청	저작권 및 라이선스 규정*
기타	특정 국가 법제도 적용 명시*
	기타
접근 통제	
오픈배고 적용	오픈배고 적용
모든 데이터에 다운로드 목적으로 접근하려면 등록 필수	모든 데이터에 다운로드 목적으로 접근하려면 등록 필수
일부 데이터에 다운로드 목적으로 접근하려면 등록 필수	일부 데이터에 다운로드 목적으로 접근하려면 등록 필수
모든 데이터에 접근하려면 공식적인 신청서 제출	모든 데이터에 접근하려면 공식적인 신청서 제출
일부 데이터에 접근하려면 공식적인 신청서 제출	일부 데이터에 접근하려면 공식적인 신청서 제출
데이터 접근 시 시간 제한 [†]	데이터 접근 위원회(DAC) 운영*
	접근 또는 저장을 위한 요금 존재*
	부적절한 데이터 활용 시 접근 중지(termination)*
IP 범위 제한 [†]	데이터 접근수준 명시*

[†] : 선행연구에서 제시되었으나 최종 분석기준에서는 제외됨.

* : 분석 과정에서 확인되어 최종 분석기준에 추가됨.

** : 기존의 분석 기준에서 변경됨.

출처: Eschenfelder, K. R., & Johnson, A. (2014). Managing the data commons: Controlled sharing of scholarly data. *Journal of the Association for Information Science and Technology*, 65(9), p. 1761, 재수정.

행하는 동안 명시적으로 드러나지 않는 분석 기준은 제외되었으며 선행연구의 분석 기준에는 없지만 새롭게 드러난 내용일 경우 추가하거나 기존의 분석 기준을 변경하는 방식으로 분석을 진행하였다.

그 결과 최종 분석 기준은 <표 3>의 오른쪽 열에 제시된 대로 이용 통제 영역에서 12개, 접근 통제 영역에서 9개로 제시되었다. 새롭게 제시된 분석 기준은 이용 통제 영역에서 3가지로 1) 면책조항(disclaimer); 2) 저작권 및 라이선스 규정; 3) 특정 국가 법제도 적용 명시이다. 또한 변경된 기준으로 ‘학술적 목적으로 제한’하는 경우보다는 학술적 목적을 포함한 비상업적 목적으로 이용을 제한하는 경우가 많아 분석 기준을 ‘비상업적 목적으로 제한’으로 변경하였다. 접근 통제 영역에서는 1) 데이터 접근 위원회(DAC) 운영; 2) 접근 또는 저장을 위한 요금 존재; 3) 부적절한 데이터 활용 시 접근 중지(termination); 4) 데이터 접근 수준 명시 등 4가지 기준이 추가적으로 제시되었다.

명시적으로 드러나지 않아 최종 분석 기준에서 제외된 것으로 이용 통제 영역에서는 1) 데이터의 안전한 저장 요청; 2) 잘못된 결과를 초래할 수 있는 데이터 이용 금지이었으며, 접근 통제 영역에서는 1) 데이터 접근 시 시간 제한; 2) IP 범위 제한인 것으로 나타났다. 이 중에서 잘못된 결과를 초래할 수 있는 데이터 이용 금지는 부적절한 데이터 활용 시 데이터로의 접근을 중지하는 것과 관련되어 있는데 실제 정책에서는 데이터 접근 중지(termination)로 제시하는 경우가 대부분이었으므로 이를 접근 통제 영역의 최종 분석 기준으로 제시하였다.

본 연구에서는 <표 3>에 제시된 최종 분석 기준에 따라 각 레포지터리의 접근 및 이용정책 또는 관련 내용을 분석하였으며 레포지터리 별로 분석 기준에서 제시된 내용이 명시되어 있는 경우 ●으로 표시하였다. 또한 각 분석 기준이 조사대상 레포지터리 정책에서 제시되어 있는 빈도를 조사하여 그 빈도수에 따라 분석 기준을 정렬하였다. 분석 결과는 생명과학 및 보건과학 분야와 화학, 지구환경과학 및 물리학 분야, 그리고 사회과학 및 일반과학으로 나누어 제시하였다.

가. 생명과학 및 보건과학 분야 데이터 레포지터리 접근 및 이용 정책

생명과학 및 보건과학 분야의 조사대상 레포지터리는 모두 20개로 전체 조사대상의 54%를 차지하고 있으며 NIH를 중심으로 유전체 데이터의 공유가 활발히 논의되고 있는 시점에서 데이터의 접근과 이용 통제에 대한 내용을 다양한 범위에서 규정하고 있었다. 이는 <표 4>에 나타난 것과 같이 조사 대상 레포지터리에 공통된 정책 내용이 존재한다기보다는 다양한 내용이 분산적으로 제시되고 있다는 점에서 확인할 수 있었다.

20개의 레포지터리 중에서 절반 이하의 빈도수이기는 하지만 이용통제 영역 중 8개의 레포지터리에서는 면책 조항(disclaimer)이 공통적인 정책 내용으로 나타났으며 6개의 레포지터리에서 데이터 재이용시 비상업적 목적으로 제한할 것을 명시하고 있었다. 5개의 레포지터리에서는 인용 등 데이터 생산자에 대한 기여를 인정할 것을 요청하고 있었다. 아울러 저작권

및 라이선스 규정을 명시하고 있는 레포지터리도 5곳인 것으로 나타났다. 이 외에도 3-4개의 레포지터리에서는 데이터에 나타난 개인의 비밀유지와 승인되지 않은 개인에 의한 데이터 배포는 금지한다는 내용을 제시하고 있었다.

Disclaimer

Browsing and using the data in NeuroMorpho.Org is at the user's own risk. George Mason University, the Krasnow Institute for Advanced Study, their leaders, and employees make no representations, warranties or guarantees of any kind or nature regarding the content or accuracy of the information accessible on NeuroMorpho.Org.

To the fullest extent permissible pursuant to applicable federal, state and local laws, NeuroMorpho.Org, its developers, and any party involved in its creation, maintenance, or modification, disclaim and shall not be liable for any damages, including, but not limited to, direct, indirect, incidental, consequential, punitive damages, attorneys' fees or costs arising out of or related to access to or use of NeuroMorpho.Org. Links are provided as a convenience only, and NeuroMorpho.Org makes no representations, warranties or guarantees regarding their contents.

Please email the NeuroMorpho.Org Administrator at nmadmin@gmu.edu or neuromorpho@gmail.com for any clarifications, questions, suggestions, and comments. Thank you for your support and cooperation.

<그림 1> NeuroMorpho.org의 이용 약관(Terms of Use) 중 면책조항
출처: <http://neuromorpho.org/useterm.jsp> [cited 2016.8.8].

면책조항의 예로서 신경과학 분야 데이터 레포지터리인 NeuroMorpho.org의 면책 조항을 <그림 1>에 제시하였다. 데이터의 활용에 대한 책임은 이용자에게 있으며 데이터 레포지터리 및 관련 기관들은 레포지터리에서 제공하는 콘텐츠와 정보 그리고 링크로 연결된 내용의 정확성을 보장하지 않는다는 것을 명시하고 있다. 또한 해당 데이터 레포지터리의 접근과 이용으로 인해 발생한 어떠한 손해에 대해서도 데이터 레포지터리와 관련 기관들은 법적, 금전적인 책임이 없음을 주장하고 있다.

비상업적인 목적의 데이터 활용을 명시하고 있는 6개의 레포지터리 중에서도 신경과학 분야의 FCP/INDI만 비상업적 목적의 활용을 허용함을 규정하고 있으며 나머지 5개 레포지터리는 모두 학술적 목적의 데이터 재이용을 강조하고 있어 학술적 목적으로 제한하는 것이 일반적임을 알 수 있다.

인용 등 데이터 생산자에 대한 기여를 인정하는 것은 5개의 레포지터리에서 강조하고 있는데 주로 레포지터리에서 제공하는 데이터에 대한 인용 형식을 제시하거나 실제 데이터를 인용한 출판물의 링크를 제공하고 있다. 예를 들어 생명과학 커뮤니티 데이터 레포지터리인 Xenbase의 경우 'Citing Xenbase'라는 링크를 통해 Xenbase의 데이터를 인용한 출판물 리

스트를 제시하고 있다.

저작권 또는 라이선스에 대한 규정을 제공하고 있는 레포지터리는 모두 5개이었으며 이중 4개 레포지터리에서 크리에이티브 커먼스(Creative Commons, CC) 라이선스를 활용하고 있는데 각각 다른 종류의 CC 라이선스를 명시하고 있다. 즉, 오믹스 분야 DIP에서는 CC 저작자 표시-변경 금지 3.0 [CC Attribution-No Derivative Works 3.0 (CC BY-ND 3.0)] 라이선스를, 수학적 모델링 분야 BioModels Database에서는 CCzero (CC0) 라이선스를, KiMoSys에서는 CC 저작자 표시 4.0 국제 [CC Attribution 4.0 International (CC BY 4.0)] 라이선스를 활용하고 있다. 또한 생명과학 커뮤니티 분야의 Xenbase는 CC 저작자 표시 3.0 [CC Attribution 3.0 (CC BY 3.0)] 라이선스를 명시하고 있다.

CC BY-ND 3.0 라이선스의 경우 저작자 표시(Attribution)와 변경 금지(No Derivative Works)를 준수하면 상업적 목적의 데이터 재이용도 허용된다. 또한 CC0 라이선스를 활용한다는 의미는 데이터를 공공 영역(public domain)으로 양도한 것으로 저작권자의 허락 없이 누구나 자유롭게 데이터를 재이용할 수 있다는 뜻이다. CC BY 4.0이나 CC BY 3.0은 저작자 표시 즉 데이터에 대한 인용만 제공한다면 데이터 재이용에 제한을 두지 않음을 의미한다. 저작권 및 라이선스 규정을 제시하는 5개 레포지터리 중 2개 레포지터리에서는 저작권법 혹은 지적재산권법을 언급하고 있었는데 보건과학 분야 VSD에서는 국제저작권법 및 지적재산권 법률에 의거하여 VSD의 모든 콘텐츠가 보호받음을 제시하였다. Xenbase의 경우 레포지터리 명칭인 'Xenbase'가 등록된 상표(registered trademark)이고, 레포지터리에 저장되어 있는 이미지 등은 저작권의 보호를 받는 저작물임을 명시하고 있다.

개인의 비밀 유지를 강조하고 있는 사례로서 보건과학분야 NDAR(National Database of Autism Research) 레포지터리의 경우 자폐증 환자의 유전형 데이터와 표현형 데이터와 같은 민감한 데이터를 포함하고 있으므로 자폐증과 관련된 연구 참여자들의 개인정보 보호와 비밀유지를 매우 중시하고 있다. 따라서 연구 참여자들이 데이터 공유에 동의한다는 내용을 포함하는 사전 동의서(informed consent)를 제출할 것을 제시하고 있다. 또한 NDAR 레포지터리에 제출되는 데이터의 경우 특정 개인이나 가족 등을 식별할 수 있는 정보를 모두 삭제해야 한다. 뿐만 아니라 NDAR의 지원 기관인 NIH에서는 법적인 요구 등 강제적인 요청에 의한 개인정보 공개에 대응하기 위해서 비밀유지 인증(Certificate of Confidentiality, CoC) 제도를 운영하고 있다. CoC는 연구과제가 진행되는 기관을 대상으로 발행되는데 법적인 요구 등 강제성 있는 요청 하에서도 연구 참여자들의 개인 정보를 공개할 수 없도록 함으로써 연구자들을 보호하는 제도이다. 다만 이 인증은 연구과제의 시작에서부터 종료 시점까지 유효하며 연구 참여자들이 자발적으로 공개를 원할 경우에는 인증이 적용되지 않는다는 제한점이 있다(NIH 2016).

〈표 4〉 생명과학 및 보건과학 레포지터리 접근 및 이용 정책 분석 결과

학문분야		핵심 서열		분자/조분자 구조		신경과학		생명과학				보건과학				계				
세부 영역	레포지터리명	ENA	PCDD8EMDB	Neuro Morpho INDI .org	FCP/ Array Express	db GAP	EGA	DIP	Proteome Xchange	Bio Models DB	KiMo Sys	EuPath DB	Vector Base	Xen base	Bio GRID		NAHDAP	NDAR	VSD	NDCT
이용 통제																				
분석기준	면적조항 (disclaimer)		●	●	●	●		●	●	●	●	●	●			●		●		8
	비상업적 목적으로 제한				●	●				●	●	●		●				●		6
	인용을 인정하게			●				●				●	●	●	●					5
	저작권/라이선스 규정 존재							●	●	●	●				●			●		5
	데이터에 나타난 개인의 비밀유지				●	●										●		●		4
	송인되지 않은 개인에 의한 데이터				●	●			●											3
	배포 금지																	●		2
	송인된 연구가 아니면 데이터 배포 금지						●										●			1
	데이터 이용에 대해 레포지터리나 데이터 생산자에게 보고 요청					●														1
	데이터의 이용 목적 혹은 특정 시기가 지나면 데이터 폐기					●														1
데이터 관련 보안 문제 발생시 보고 요청						●												●		1
특정 국가 법제도 적용			●		●	●							●	●			●			5
접근 통제																				
분석기준	열람과 적용	●	●		●	●	●	●	●			●	●	●						7
	데이터접근위원회 (Data Access Committee) 운영				●	●	●											●		5
	일부 데이터에 접근하려면 공식적인 신청서 제출				●	●												●		3
	모든 데이터에 다운로드 목적으로 접근하려면 등록 필수							●												2
	데이터 접근수준 명시	●							●							●				2
	부적절한 데이터 활용시 접근 금지																	●		2
	모든 데이터에 접근하려면 공식적인 신청서 제출							●												1
	일부 데이터에 다운로드 목적으로 접근하려면 등록 필수																			0
	접근 또는 저장을 위한 접근 통제																			0

접근 통제 영역과 관련하여 <표 4>에서와 같이 7개의 레포지터리에서 데이터 공개를 연기할 수 있게 하는 제도인 엠바고(embargo) 적용을 명시하고 있다. 이와 더불어 5개 레포지터리에서는 데이터 접근 위원회의 운영을 언급하고 있으며, 데이터 접근 위원회에 신청서를 제출하여 승인을 받은 이용자들만 데이터의 재이용이 가능함을 명시하고 있다.

엠바고의 경우 기간을 구체적으로 명시하는 레포지터리는 7개 중 2개로 분자/초분자 구조 분야의 EMDB와 생명과학 커뮤니티 분야의 VectorBase 레포지터리인 것으로 나타났다. EMDB의 경우 데이터를 레포지터리에 제출한 일자로부터 1년 이내에 공개할 것을 제안하고 있으며 VectorBase에서는 데이터 생산 후 1년 이내 혹은 논문 출판 후 1개월 이내의 기간 중에서 먼저 도래하는 일자를 기준으로 할 것을 명시하고 있다. 나머지 5개의 레포지터리에서는 엠바고를 허용하지만 구체적인 엠바고 기간을 규정하지 않고 있다. 예를 들어 오믹스 분야의 dbGAP에서는 엠바고 기간 전에 출판물이나 데이터를 공개하면 안 된다는 규정을 제시하고 있으며 핵산서열 분야의 ENA에서는 비밀유지가 필요한 데이터는 지정된 배포일 이전에 공개해서는 안 된다고 규정하고 있다.

데이터 접근 위원회의 운영은 접근 통제 영역에서 5개 레포지터리에서 제시하고 있는 내용이다. 이는 본 연구의 조사 대상 중 생명과학 및 보건과학 분야의 레포지터리에서만 나타나는 접근 통제 방법이다. Paltoo et al.(2014)에 따르면 데이터 접근 위원회는 데이터를 2차적으로 이용하고자 하는 연구자가 제출한 데이터 접근 요청서(Data Access Requests)를 심의하고 승인 여부를 결정한다. 데이터 접근 요청서에는 데이터 재이용 계획과 협동연구자 및 연구책임자가 속한 기관의 IT 담당관의 이름 및 연구책임자와 기관의 공식 서명이 포함된다. 이러한 과정을 거쳐 승인된 데이터 이용 인증서(Data Use Certification)는 일종의 협약서로서 데이터 접근 협약서(Data Access Agreement)라고 지칭되기도 하는데 특정 데이터세트의 재이용과 관련된 약관 및 조건을 규정하고 있다.

본 연구의 조사 대상 중 가장 상세한 접근 및 이용 정책을 제시하고 있는 레포지터리는 오믹스 분야의 ArrayExpress와 dbGAP인 것으로 나타났다. <표 4>에 제시된 대로 해당 레포지터리에서 제시되는 정책에서는 다양한 내용을 포괄하고 있다. 예를 들어 ArrayExpress에서는 데이터 접근 협약서의 템플릿을 제공하고 있는데 이 템플릿에서 다양한 이용 및 접근 정책 내용을 명시하고 있다. 분석 기준으로 제시된 내용 이외에 기타 내용으로 ArrayExpress의 협약서에는 데이터 재이용을 승인받은 연구자의 신상 변경 시 30일 이내에 데이터 생산자에게 보고해야 하며 연구 프로토콜 변경 전에도 데이터 생산자에게 이를 보고해야 한다는 내용이 명시되어 있다. 또한 협약서에 명시된 내용을 위반한 사실을 인지하는 즉시 데이터 생산자에게 보고해야 하며 데이터 생산자가 협약 종료료를 요청할 때는 모든 복사본, 백업카피 등 소유하고 있는 데이터를 폐기해야 한다. 이와 더불어 데이터 생산자가 협약 내용 변경을 요청할

경우 받아들일 것을 요청하고 있다. 이를 통해 데이터 재이용에 대하여 데이터 생산자의 통제 권한을 최대한으로 인정하고 있음을 알 수 있다.


나. 화학, 지구환경과학 및 물리학 분야 데이터 레포지터리 접근 및 이용 정책

조사 대상 중 화학, 지구환경과학 및 물리학 분야 레포지터리는 10개(27.0%)이었으며 이 중 8개가 지구환경과학 분야 레포지터리로 대다수가 이에 해당함을 알 수 있다(<표 5> 참조). 생명과학 및 보건과학 분야 정책에 비해 이들 학문분야 데이터 레포지터리의 접근 및 이용 정책은 몇몇 분석 기준에 집중되는 경향을 보였다.

이용 통제 영역에서는 10개 레포지터리 중 8개 레포지터리에서 저작권 및 라이선스 규정을 명시하고 있어 대다수가 이를 정책에 포함시키고 있음을 알 수 있다. 이 중에서 5개 레포지터리에서 CC 라이선스를 활용하는 것을 명시하고 있다. 지구환경과학 분야 PANGAEA와 물리학 분야 NoMaD는 CC BY 3.0 라이선스를, 지구환경과학 분야 AADC에서도 CC BY 라이선스 사용을 권장하고 있어 데이터의 저작자 표기를 강조하고 있다. 지구환경과학 분야 Seanoe의 경우 특정 CC 라이선스를 정하지 않고 적절한 CC 라이선스를 선택해서 사용할 것을 언급하고 있다. 반면에 MGDS에서는 CC 저작자표시-비영리-동일조건변경허락 4.0 [CC Attribution-NonCommercial-ShareAlike (CC BY-NC-SA 4.0)] 라이선스를 이용하도록 규정하고 있다. 이 라이선스 하에서는 저작자를 표시하면 이용이 가능하며 저작물의 변경도 가능하지만, 상업적 목적으로 데이터를 재이용할 수 없고 데이터 재이용을 통해 생산된 저작물에는 원 저작물과 동일한 라이선스를 적용해야 한다.

CC 라이선스 이외에 저작권 등 지적재산권에 대한 내용을 제시하는 경우로 화학 분야 ChEMBL은 데이터에 대한 모든 권한에 대한 내용은 데이터 접근 협약서에 명시할 것을 언급하고 있다. 지구환경과학 분야 CARD에서도 레포지터리에서 다운받아 활용한 모든 이미지 및 텍스트 자료의 저작권은 특별한 언급이 없는 한 해당 레포지터리에 있음을 규정하고 있다. 이 외에도 지구환경과학 분야 AADC의 경우 CC 라이선스의 활용을 언급함과 동시에 레포지터리의 모든 콘텐츠의 저작권은 호주 연방(Commonwealth of Australia)에 있으며 그러한 언급이 없는 경우에는 저작자에게 권한이 있음을 명시하고 있다.

이용 통제 영역에서 저작권 및 라이선스 규정 이외에도 인용 등 인정체계에 대한 요청이 6개 레포지터리에서 제시되었고 면책 조항은 4개 레포지터리에서 제시되고 있는 것으로 나타났다. 인용을 요청하는 레포지터리에서는 데이터 인용의 형식을 제공하는 등 레포지터리에 저장된 데이터를 재이용할 경우 반드시 이를 인용할 것을 권장하고 있다. 예를 들어 <그림 2>는 지구환경과학 분야의 CARD(Cold and Arid Regions Science Data Centers) 레포지터리에서 제공하는 인용 관련 정책 내용을 보여주고 있다.



Citing Data

We request that users of CARD data cite the use of our data in their work.

Citation of data acquired from CARD acknowledges data contributors, and allows us to track the use and impact of our data, to more effectively report data distribution activity to funding agencies, and to assist other users who may contact us with questions about data that are referenced in publications.

Please note that we prefer that users include a citation in the references section of the publication. While we appreciate acknowledgement within the text of the publication, we are better able to track the use of our data sets if citations are included as references.

Appropriate citations are displayed on the catalog pages for many of our data sets. If you would like to cite data for which a citation is not provided, please contact us for help.

Further, we request that users send us one reprint of any published work for our library collection. This helps us to stay aware of how our data are being used by the user community and helps us to build our library collection.

〈그림 2〉 CARD 레포지터리의 데이터 인용 관련 정책

출처: <http://card.westgis.ac.cn/about/copyright> [cited 2016.8.8].

위의 그림에서 제시된 바와 같이 CARD 레포지터리에서는 데이터 기탁자의 기여를 인정하고 데이터의 활용과 영향력을 확인할 수 있는 등 인용의 유용성을 설명하면서 데이터를 참고 문헌 내에서 인용할 것을 요청하고 있다. 또한 이용자에게 데이터를 인용하여 출판한 논문의 별쇄본(reprint)을 레포지터리에 보내줄 것을 요청하고 있어 해당 레포지터리에서 데이터의 활용도를 확인하고 추적하는데 많은 관심을 기울이고 있음을 알 수 있다.

이와 더불어 데이터 활용과 관련된 법적인 논의가 필요할 때 특정 국가의 법제도 적용을 명시하는 레포지터리도 2곳인 것으로 나타났다. 지구환경과학 분야의 Seanoe는 프랑스 법률, 물리학 분야 NoMaD는 독일법을 적용한다는 것을 언급하였다. 비상업적 목적의 데이터 활용을 명시한 레포지터리도 2곳이었는데 지구환경분야의 LTER과 MGDS이었으며 MGDS는 CC 라이선스를 통해 이를 제시하고 있었다.

이 외에도 기타 내용으로 지구환경과학 분야의 MGDS에서는 상업적 성격을 지닌 데이터(proprietary data)의 재이용에 있어서 기탁자와 이용자 간 협약서를 작성할 것을 제안하고 있다. GES DISC는 NASA(National Aeronautics and Space Administration)로부터 재정 지원을 받는 레포지터리이므로 NASA의 데이터 정책을 따른다는 것을 명시하고 있다.

접근 통제 영역에서는 엠바고 적용과 관련된 정책 내용을 언급한 레포지터리 3곳 중 구체적인 기간을 명시한 레포지터리는 2곳이었다. NERC에서는 데이터 생성 후 2년 이내에는 공개할 것을 제시하고 있고 LTER의 경우 공개하기로 결정한 데이터는 데이터 수집 후 2년 이내, 혹은 데이터 분석을 통해 주요 결과를 논문으로 출판한 시기에 공개할 것을 명시하고 있다. 구체적인 엠바고 기간을 명시하지 않은 PANGAEA에서는 특정 연구과제에서 엠바고를 명시하고 있는 경우 그 과제에서 생산된 데이터는 해당 규정을 따를 것을 제안하고 있다.

〈표 5〉 화학, 지구환경과학 및 물리학 분야 레포지터리 접근 및 이용 정책 분석 결과

학문분야	화학/ 화학생명 과학	지구환경과학										물리, 천체물리 및 천문학		계
		레포지터리명	ChEMBL	CARD	NERC	LTER	PANGAEA	AADC	MGDS	Seano	GES	DJSC	NoMaD	
분석기준														
이용 통제														
저작권/라이선스 규정 존재	●	●	●	●	●	●	●	●	●	●	●	●	●	8
인용 등 인정체계(acknowledgement)요청		●	●	●	●	●	●	●	●	●	●	●	●	6
면책조항 (disclaimer)	●	●	●				●					●	●	4
특정 국가 법제도 적용									●			●	●	2
비상업적 목적으로 제한					●				●					2
데이터에 나타난 개인의 비밀유지	●													1
승인된 연구가 아니면 데이터 배포 금지			●											1
데이터 이용에 관해 레포지터리나 데이터 생산자에게 보고 요청							●							1
승인되지 않은 개인에 의한 데이터 배포 금지								●						0
데이터의 이용 후 혹은 특정 시기가 지나면 데이터 폐기														0
데이터 관련 보안 문제 발생시 보고 요청														0
기타								●		●	●	●		4
접근 통제														
웹보고 적용				●	●	●		●						0
데이터 접근수준 명시					●									3
접근 또는 저장을 위한 요금 존재					●									1
데이터접근위원회(Data Access Committee) 운영														1
일부 데이터에 접근하려면 공식적인 신청서 제출														0
모든 데이터에 다운로드 목적으로 접근하려면 등록 필수														0
일부 데이터에 다운로드 목적으로 접근하려면 등록 필수														0
모든 데이터에 접근하려면 공식적인 신청서 제출														0
부적절한 데이터 활용시 접근 중지(termination)														0

데이터 접근 수준을 명시하는 레포지터리는 지구환경과학 분야 LTER 1곳인데 누구나 접근 및 재이용할 수 있는 데이터를 “Type I” 데이터로, 접근 제한을 두는 데이터를 “Type II” 데이터로 구분하여 제시하고 있다. Type II 데이터의 경우 데이터 생산자가 제시하는 조건에 의해 접근 및 이용이 제한되며 그 근거로써 비밀유지, 상업적 이해관계, 저작권 보호 또는 희귀 및 멸종위기종의 서식지 보호 등을 명시하고 있다.

데이터의 접근 및 저장을 위해 비용을 요청하는 레포지터리는 지구환경과학 분야 NERC (National Environmental Research Council)의 데이터 센터 1곳이었다. 여기에서는 특정한 경우에 한하여 데이터 기탁자에게 데이터 관리에 대한 비용을 청구한다는 내용을 제시하고 있다. 그러한 경우로는 (1) 대용량이며 복잡한 구조를 가진 데이터의 기탁; (2) 데이터에 대한 제 3자 라이선스 조건에서 무료 배포를 금지하거나 레포지터리에서 비용을 부담할 것을 요청하는 경우이다.

생명과학 및 보건과학 레포지터리와 비교해보았을 때 인용 관련 내용과 면책 조항은 공통적으로 명시되는 빈도수가 높은 내용임을 알 수 있다. 그러나 생명과학 및 보건과학 레포지터리에서는 비상업적 또는 학술적 목적의 재이용을 명시하는 경우가 많았던 반면 화학·지구환경과학·물리학 분야 레포지터리에서는 그러한 언급은 레포지터리 2곳에서만 제시되고 있었다. 뿐만 아니라 생명과학 및 보건과학 레포지터리에서는 나타나지 않았던 내용이지만 데이터 기탁에 대한 비용을 청구하는 경우를 명시한 사례와 법적인 문제가 발생할 경우 적용되는 국가법을 언급한 사례들이 있었다.

다. 사회과학 및 일반과학 분야 데이터 레포지터리 접근 및 이용 정책

사회과학 분야로 분류된 데이터 레포지터리 5곳과 일반과학 분야 데이터 레포지터리 2곳 등 모두 7개 레포지터리를 대상으로 분석한 결과 이용 통제 영역에서 가장 많은 수인 5개 레포지터리에서 저작권 및 라이선스 규정을 명시하고 있었고 모두 CC 라이선스를 언급하고 있었다. Harvard Dataverse와 figshare는 CC0 라이선스를 제시하여 데이터를 공공 영역으로 공개하도록 하고 있고, OpenICPSR과 Dryad는 각각 CC BY 4.0과 CC BY 3.0을 명시하고 있어 데이터의 인용을 강조하고 있음을 알 수 있다. UK Data Service-Reshare의 경우는 CC 라이선스 또는 Open Government License (OGL) 둘 중의 하나를 사용할 것을 권장하고 있다. OGL은 공공영역에 속한 정보(public sector information)를 위한 라이선스로 영국국립기록관(The National Archives)에서 관리하고 있다. 이 라이선스 하에서는 정보의 복사, 출판, 배포, 전송 및 변경이 가능하며 상업적 또는 비상업적 목적으로의 이용이 모두 허용된다. 다만 해당 정보의 출처 등 저작자에 대한 인용이 반드시 요청된다. 내용적으로는 CC 저작자표시 라이선스와 동일하므로 현재 최신 버전인 OGL 3.0과 CC BY 4.0은 상호 호

환 가능한 라이선스로 규정하고 있다(OGI 2016).

면책 조항은 4개 레포지터리에서, 인용에 대한 요청은 3개 레포지터리에서 제시하고 있어 이들은 앞서 조사한 생명과학 및 보건과학 분야 레포지터리 또는 화학, 지구환경과학 및 물리학 분야 레포지터리와 공통적으로 자주 등장하는 정책임을 알 수 있다. 이 외에도 데이터에 나타난 개인의 비밀 유지와 관련하여 3개의 레포지터리에서 이를 언급하고 있었다. 예를 들어 Harvard Dataverse 정책에서는 개인을 식별할 수 있는 정보를 제외하고 데이터를 기탁할 것을 명시하고 있다. 그러한 정보의 예로 사회보장번호(Social Security Number), 신용카드번호, 의료기록번호, 보험 번호(health plan number), 계좌번호, 그리고 지문, 홍채, DNA 등 생체인식정보 등을 제시하고 있다.

〈표 6〉 사회과학 및 일반과학 분야 레포지터리 접근 및 이용 정책 분석 결과

분석기준	학문분야	사회과학					일반과학		계
	레포지터리명	Harvard Dataverse	Open ICPSR	UK Data Service - ReShare	OSF	ADS	Dryad Digital Repository	figshare	
이용 통제									
저작권/라이선스 규정 존재		●	●	●			●	●	5
법적책임 없음 / 면책조항 (disclaimer)		●			●		●	●	4
인용 등 인정체계(acknowledgement)요청		●		●		●			3
데이터에 나타난 개인의 비밀유지		●		●		●			3
비상업적 목적으로 제한				●					1
특정 국가 법제도 적용 (프랑스, 스웨덴, 독일)									0
승인되지 않은 개인에 의한 데이터 배포 금지									0
승인된 연구가 아니면 데이터 배포 금지									0
데이터 이용에 관해 레포지터리나 데이터 생산자에게 보고 요청									0
데이터의 이용 후 혹은 특정 시기가 지나면 데이터 폐기									0
데이터 관련 보안 문제 발생시 보고 요청									0
기타									0
접근 통제									
부적절한 데이터 활용시 접근 중지(termination)		●		●	●		●	●	5
접근 또는 저장을 위한 요금 존재			●				●	●	3
오픈방과 적용						●	●		2
데이터 접근수준 명시				●	●				2
일부 데이터에 다운로드 목적으로 접근하려면 등록 필수				●					1
데이터접근위원회(Data Access Committee) 운영									0
일부 데이터에 접근하려면 공식적인 신청서 제출									0
모든 데이터에 다운로드 목적으로 접근하려면 등록 필수									0
모든 데이터에 접근하려면 공식적인 신청서 제출									0

접근 통제 영역에서는 5개 레포지터리에서 부적절한 데이터 활용이 발견될 경우 데이터 이용자의 접근을 중지한다는 것을 규정하고 있었다. 예를 들어 사회과학분야 Center for Open Science(COS)의 OSF(Open Science Framework) 레포지터리에서는 <그림 3>과 같은 접근 중지 규정을 제시하고 있다. 이용자가 데이터 이용과 관련된 법을 위반하거나 관련

기관 및 개인의 이익을 해하는 경우 또는 이용 약관을 위반하는 경우에 레포지터리를 운영하는 기관인 COS에서 데이터로의 접근을 중지하거나 제한하는 권한을 행사함을 명시하고 있다.

14. TERMINATION

These Terms of Use are effective until terminated by either party. If you no longer agree to be bound by these Terms, you must cease use of the Websites and Services. If you're dissatisfied with the Websites, their content, our Services or any of these terms, conditions, and policies, your sole legal remedy is to discontinue using the Websites and Services. The COS reserves the right to terminate or suspend your access to and use of the Websites and Services, or parts of the Websites and Services, without notice, if we believe, in our sole discretion, that such use is (i) in violation of any applicable law; (ii) is harmful to our interests or the interests, including intellectual property or other rights, of another person or entity; or (iii) where the COS has reason to believe that you are in violation of these Terms of Use. Termination for any reason shall not terminate any of your warranties or any of your indemnification and other obligations under these Terms of Use, or any licenses that you have license, use or other rights that you may have granted in connection with any of your contributions.

〈그림 3〉 OSF의 접근 중지 규정

출처: https://github.com/CenterForOpenScience/centerforopenscience.org/blob/master/TERMS_OF_USE.md [cited 2016.8.8].

접근 또는 저장을 위한 요금에 대해서도 3개 레포지터리가 이를 명시하고 있는데 이 중 대표적인 사례로 OpenICPSR이 있다. 방대한 사회과학 데이터 아카이브로 유명한 ICPSR (Interuniversity Consortium for Political and Social Research)에서 별도의 데이터 레포지터리로 OpenICPSR을 운영하고 있다. 여기에서는 연방정부의 지원을 받은 사회과학 및 행동과학 분야 연구과제에서 생성된 데이터를 기탁받고 있다. 이러한 데이터는 공적자금으로 수행된 연구의 데이터이므로 데이터에 대한 접근 공개 의무를 준수해야 하며 OpenICPSR이 이를 지원하고 있다.

OpenICPSR은 기존의 ICPSR 회원에게는 무료 기탁 서비스를 제공하지만 비회원일 경우 유료 서비스를 제공하며 다음의 3단계 서비스 체계를 갖추고 있다. 첫째, Self-Deposit Package의 경우 ICPSR회원일 경우는 무료이지만 비회원은 연간 600불을 지불해야 한다. 데이터에 DOI가 부여되며 레포지터리에서는 데이터 기탁자가 제공한 메타데이터를 검토한다. 둘째, Professional Curation Package로 ICPSR에서 제공하는 데이터 큐레이션 서비스를 이용할 수 있고 데이터의 복잡성 정도에 따라 요금이 결정된다. 셋째, OpenICPSR for Institutions and Journals로 기관이나 학술지를 대상으로 데이터 레포지터리 유료 서비스를 제공한다.

반면에 일반과학 분야의 Dryad와 figshare는 무료 서비스를 지향하지만 특수한 경우 유료 서비스를 제공하고 있음을 명시하고 있다. Dryad 레포지터리에서는 20GB를 초과하는 데이터의 경우 저장을 위한 추가 요금을 납부할 것을 규정하고 있다. figshare에서는 기본 서비스(basic service)는 무료이나 프리미엄 서비스(premium service)는 유료로 운영됨을 명시하고 있다.

엠바고 적용에 대해서는 2개 레포지터리에서 명시하고 있는데 사회과학분야 ADS의 경우 데이터의 민감성 정도에 따라 엠바고 기간이 달라질 수 있음을 언급하고 있으나 구체적인 기간은 명시하지 않았다. 그에 비해 Dryad 레포지터리에서는 최초로 학술지 논문을 출판한 이후 1년 이내에 데이터를 공개할 것을 규정하고 있다.

데이터 접근 수준에 있어서 2개 레포지터리가 이를 명시하고 있는데 UKDS-Reshare 레포지터리에서는 오픈 데이터(Open Data)와 접근 통제된 데이터(Safeguarded and Controlled Data)로 접근 수준을 구분하고 있다. 오픈 데이터는 OGL 또는 CC 라이선스에 기반을 두고 재이용할 수 있는 데이터로 UKDS에 이용자 등록 없이도 접근 가능하다. 그러나 접근 통제된 데이터를 이용하기 위해서는 UKDS에 이용자 등록이 필수적이며 등록 시 이용약관인 End User License에 동의하여야 한다. 또한 OSF에서는 누구나 접근이 가능한 “Public” 콘텐츠와 레포지터리 관리자 또는 관리자가 접근을 허용한 이용자만이 접근 가능한 “Private” 콘텐츠로 접근 수준을 제시하고 있다.

3. 분석 결과 요약

본 연구에서는 총 37개의 데이터 레포지터리의 접근 및 이용 정책을 분석하는데 있어 분야별로 생명과학 및 보건과학 분야 20개, 화학, 지구환경과학 및 물리학 분야 10개, 사회과학 및 일반과학 7개의 레포지터리로 나누어 정책 내용 분석을 실시하였다. 분야를 막론하고 과반수에 가까운 데이터 레포지터리에서 공통적으로 제시한 정책 내용은 이용 통제 영역에서는 ‘면책조항’과 ‘인용 등 인정체계 요청’, ‘저작권 및 라이선스 규정’이었으며 접근 통제 영역에서는 ‘엠바고 적용’인 것으로 나타났다.

면책 조항에서는 레포지터리에서 제공하는 데이터 자체의 품질 및 데이터 재이용을 통해 생산되는 2차적 정보의 정확성 및 신뢰성에 대하여 데이터를 생산하고 기탁한 연구자 및 데이터를 2차적으로 활용하는 이용자의 법적 책임을 강조하고 있다. 이는 데이터의 접근 및 활용에서 발생할 수 있는 다양한 법적 분쟁에서 레포지터리의 책임을 최소화하기 위한 장치임을 알 수 있다.

인용을 통해 데이터 생산자의 기여를 인정해야 한다는 내용 역시 공통적으로 자주 언급되고 있으며 그러한 레포지터리마다 인용 형식을 표시하거나 예시를 제시하는 방식으로 데이터의 인용을 적극 독려하고 있었다. 이와 같은 인용의 중요성은 저작권 및 라이선스 규정을 명시하는데 있어서도 나타나는데 이러한 규정을 제시한 19개 레포지터리 중 14개 레포지터리에서 저작자표시(CC BY)를 기본으로 하는 CC 라이선스를 채택하고 있다. 다만 이 중에서 3개 레포지터리는 CC0 라이선스를 규정하고 있는데 이는 데이터를 공공영역정보로 활용하

는 라이선스이므로 저작자표시를 규정하는 것은 아니지만 인용 자체가 연구윤리의 일부이므로 여전히 인용을 중요한 요소로 인정하고 있다고 볼 수 있다(Creative Commons 2016).

엠바고 적용에 대한 언급은 접근 통제 영역에서 자주 등장하는 내용이었는데 이를 제시하고 있는 12개 레포지터리 중에서 실제 구체적인 기간을 제시하고 있는 레포지터리는 모두 5개로 생명과학 분야의 EMDB와 VectorBase, 지구환경과학 분야의 NERC와 LTER, 그리고 일반과학 분야의 Dryad 레포지터리이었다. 엠바고 기간을 제시할 때 레포지터리별로 기준일을 달리 하고 있었는데 조사 대상 레포지터리에서는 데이터 생산 혹은 수집 시점, 데이터를 레포지터리에 제출한 시점, 데이터를 기반으로 논문을 출판한 시점 등 세 가지 일자를 기준으로 하고 있었다. 데이터의 생성/수집 또는 제출 시점을 기준으로 하면 1년 내지 2년 내에 공개할 것을 규정하고 있었으며 학술지논문 출판일을 기준으로 하여 즉시 공개 또는 1개월 이내 공개를 명시하고 있었다.

분야를 막론하고 공통적으로 자주 언급되는 정책 내용 외에 특정 분야 레포지터리에서 주로 언급되는 내용도 존재하였다. 생명과학 및 보건과학 분야의 레포지터리에서는 비상업적 혹은 학술적 목적으로 데이터 재이용을 제한하는 규정의 빈도 수가 다른 분야에 비해 많았다. 또한 데이터 접근 위원회의 운영은 생명과학 및 보건과학 분야에서만 활성화되어 있는 이용 통제 정책인 것으로 나타났다. 이러한 차이점은 생명과학 및 보건과학 분야 데이터의 특성상 연구 참여자의 개인정보 보호와 비밀유지가 필수적인 경우가 많기 때문인 것으로 이해할 수 있다.

사회과학 및 일반과학 데이터 레포지터리에서는 부적절한 데이터 활용 시 데이터에 대한 접근을 중지하는 조항이 자주 언급되었다. 사회과학 분야 데이터 역시 인간을 대상으로 하여 수집되는 경우가 많으므로 개인정보 보호와 비밀유지에 관한 내용도 화학, 지구환경과학 및 물리학 분야 레포지터리에 비해서는 제시되는 빈도 수가 높았다.

정책 내용의 다양성에 있어서 생명과학 및 보건과학 분야 레포지터리에서 여러 가지 정책 내용들이 분산적으로 제시되고 있었던 반면 그 외 분야 레포지터리에서는 상대적으로 저작권/라이선스 규정, 인용 규정 또는 부적절한 데이터 활용 시 접근 중지 등 몇몇 정책 내용에 집중되는 경향을 보였다. 또한 생명과학 및 보건과학 분야에서는 나타나지 않았지만 그 외의 분야 레포지터리 중 일부에서는 데이터의 용량이나 복잡성 정도에 따라 요금을 부과하는 정책을 운영하고 있었다. 이와 같이 분야별로 개인정보 보호와 민감한 데이터의 비밀 유지, 부적절한 데이터 이용 금지, 데이터 생산자의 기여도와 권한을 인정하기 위한 방안으로서 다양한 데이터 이용 및 접근 통제 정책을 제시하고 있음을 알 수 있다.

IV. 결론 및 제언

본 연구에서는 오픈 사이언스를 구성하는 하나의 축인 오픈 데이터 실현에 있어 데이터 레포지터리의 역할이 강화되고 있는 상황에서 레포지터리에서 규정하는 데이터 접근 및 이용 정책 분석을 통해 실제로 어떠한 접근과 이용 통제가 이루어지고 있는지를 살펴보았다. 학문 분야에 상관없이 공통적으로 언급되는 정책 요소들을 통해 인용의 중요성 및 저작권이나 라이선스 혹은 엠바고 적용을 통한 권한 보호 등 데이터 활용에서 보편적으로 강조되는 내용들을 확인할 수 있었다. 그러나 분야 별로 통제 기제의 다양성에는 차이가 있는 것으로 나타났으며 분야 별로 통제 및 제한이 요청되는 영역과 그 정도에 차이가 있음을 알 수 있다.

데이터의 공유와 재이용이 학문 발전에 미치는 긍정적인 영향에 대하여 많은 연구자들이 동의하지만 그에 따른 부정적 영향에 대한 우려가 데이터 공유의 확산을 저해하는 요인이기도 하다. 특히 생명과학 및 보건과학 분야의 데이터는 본질적으로 연구 참여자의 개인정보를 포함하는 민감한 데이터인 경우가 많아 이러한 데이터를 공유하는데 따르는 부정적 영향을 최소화하기 위해서는 데이터의 이용과 접근에 대한 적절한 통제 기제가 필요하다는 의견이 지배적이다. 따라서 데이터의 개방적 활용과 더불어 지속가능한 관리와 접근을 가능하게 하는 안정된 서비스의 제공 사이에 균형을 이루는 정책적 논의가 필요하다.

이와 함께 데이터 이용 및 접근 통제를 규정하는 것이 데이터의 공유를 촉진시키는지 또는 지속가능하고 안정적인 데이터 관리에 기여하는지에 대한 실증적인 논의는 부족한 상황이며 향후 이러한 상관관계를 분석할 수 있는 연구가 필요할 것이다. 또한 데이터를 재이용하고자 하는 연구자들은 이러한 통제 기제를 어떻게 받아들이고 있는지에 대한 연구도 수행될 필요가 있다. 이용자의 입장에서 체계적으로 수립된 레포지터리 정책 및 표준, 기술적인 인프라는 레포지터리에 대한 이용자의 신뢰를 높일 수 있는 바탕이 된다. 그러나 많은 경우 접근 및 이용 통제 정책은 데이터 생산자 혹은 기탁자의 입장을 반영한 것이므로 이용자의 입장에서 이러한 정책이 가지는 의의 및 한계점을 논의하는 연구도 필요할 것이다.

참고문헌

Creative Commons. 2016. CC0. <https://creativecommons.org/share-your-work/public-domain/cc0/> [cited 2016.8.8].

- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747–1762.
- Doshi, J. A., Hendrick, F. B., Graff, J. S., & Stuart, B. C. (2016). Data, Data Everywhere, but Access Remains a Big Issue for Researchers: A Review of Access Policies for Publicly–Funded Patient–Level Health Care Data in the United States. *eGEMS*, 4(2). [cited 2016.8.8].
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4827788/>
- Eschenfelder, K. R., & Johnson, A. (2014). Managing the data commons: Controlled sharing of scholarly data. *Journal of the Association for Information Science and Technology*, 65(9), 1757–1774.
- Gitter, D. M. (2010). Application of Data Access Policies Designed for Genomewide Association Studies to Smaller Scale Databases. *The John Marshall Law School Review of Intellectual Property Law*, 10(3). [cited 2016.8.8].
<http://repository.jmls.edu/ripl/vol10/iss3/3/>
- Kaye, J., & Hawkins, N. (2014). Data sharing policy design for consortia: challenges for sustainability. *Genome medicine*, 6(1). [cited 2016.8.8].
<http://www.ncbi.nlm.nih.gov/pubmed/24475754>
- Nicol, A., Caruso, J., & Archambault, É. (2013). *Open data access policies and strategies in the european research area and beyond*. Science–Metrix.
http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf [cited 2016.8.8].
- NIH. 2016. Certificate of Confidentiality.
<https://humansubjects.nih.gov/coc/index> [cited 2016.8.8].
- OECD. 2015. Making open science a reality.
http://www.oecd-ilibrary.org/science-and-technology/making-open-science-a-reality_5jrs2f963zs1-en [cited 2016.8.8].
- Paltoo, D. N., Rodriguez, L. L., Feolo, M., Gillanders, E., Ramos, E. M., Rutter, J., ... & Caulder, M. (2014). Data use under the NIH GWAS data sharing policy and future directions. *Nature genetics*, 46(9), 934.
- Ramos, E. M., Din–Lovinescu, C., Bookman, E. B., McNeil, L. J., Baker, C. C., Godynskiy, G., ... & Starks, V. L. (2013). A mechanism for controlled access to GWAS data: experience of the GAIN Data Access Committee. *The American Journal of Human Genetics*, 92(4): 479–488.

- Pampel, H., & Dallmeier-Tiessen, S. (2014). Open research data: From vision to practice. In *Opening science* (pp. 213–224). Springer International Publishing.
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., ... & Dierolf, U. (2013). Making research data repositories visible: the re3data.org registry. *PloS one*, 8(11), e78080. [cited 2016.8.8].
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>
- Scientific Data. 2016a. Data policies. [cited 2016.8.8].
<http://www.nature.com/sdata/policies/data-policies#repo-suggest>
- Scientific Data. 2016b. For Authors. [cited 2016.8.8].
<http://www.nature.com/sdata/publish/for-authors#aims-scope>
- Shabani, M., Knoppers, B. M., & Borry, P. (2015). From the principles of genomic data sharing to the practices of data access committees. *EMBO molecular medicine*, 7(5), 507–509.
- Sydes, M. R., Johnson, A. L., Meredith, S. K., Rauchenberger, M., South, A., & Parmar, M. K. (2015). Sharing data from clinical trials: the rationale for a controlled access approach. *Trials*, 16(1). [cited 2016.8.8].
<https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-015-0604-6>
- TNA. 2016. Open Government License for public sector information. [cited 2016.8.8].
<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

〈부록 1〉 조사 대상 데이터 레포지터리 URL 및 데이터 접근 및 이용 정책 웹페이지 URL

학문분야	레포지터리	Repository URL	정책 웹페이지 URL
생물분야	ENA	http://www.ebi.ac.uk/ena/	http://www.ebi.ac.uk/ena/about/data-availability-policy
	PCDDb	http://pcddb.cryst.bbk.ac.uk/	http://pcddb.cryst.bbk.ac.uk/tandc.php
	EMDB	http://www.emdatabank.org/	http://www.emdatabank.org/hold_policy.html
	NeuroMorpho.org	http://neuromorpho.org/neuromorpho/index.jsp	http://neuromorpho.org/userterm.jsp
생명과학	FCP/NDI	http://fcon_1000.projects.nitrc.org/	http://fcon_1000.projects.nitrc.org/ndi/req_access.html
	ArrayExpress	http://www.ebi.ac.uk/arrayexpress/	http://www.ebi.ac.uk/arrayexpress/help/data_availability.html
	dbGAP	http://www.ncbi.nlm.nih.gov/gap	https://dbgap.ncbi.nlm.nih.gov/aa/GWAS_Code_of_Conduct.html
	EGA	http://www.ebi.ac.uk/ega/	https://www.ebi.ac.uk/ega/about/FAQ
	DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	http://dip.doe-mbi.ucla.edu/dip/termsOfUse.html
	ProteomeXchange	http://www.proteomeexchange.org/	http://www.proteomeexchange.org/sites/proteomeexchange.org/files/images/pxcollaborativeagreement.pdf
보건과학	BioModels Database	http://biomodels.net/	http://www.ebi.ac.uk/biomodels-main/termsOfUse
	KIMoSys	http://www.kimosys.org/	http://www.kimosys.org/documentation#17
	EuPathDB	http://eupathdb.org/eupathdb/	http://eupathdb.org/EuPathDB_datasubm_SOP.pdf
	VectorBase	http://www.vectorbase.org/index.php	https://www.vectorbase.org/data-policies
	Xenbase	http://www.xenbase.org/common/	http://www.xenbase.org/other/static/aboutXenbase.jsp
	BioGRID	http://thebiogrid.org/	http://wiki.thebiogrid.org/doku.php/terms_and_conditions
	NAHDAP	http://www.icpsr.umich.edu/icpsrweb/NAHDAP/index.jsp	http://www.icpsr.umich.edu/icpsrweb/content/NAHDAP/data/restricted.html
	NDAR	http://ndar.nih.gov/	https://ndar.nih.gov/ndarpublicweb/Documents/NDAR_Policy.pdf
	VSD	https://www.virtualskelton.ch/	https://www.smir.ch/Home/TermsAndConditions
	NDCT	http://ndct.nih.gov/	https://data.archive.nih.gov/ndct/s/SharedContent/about/policy.html
화학	CHEMBL	https://www.ebi.ac.uk/chembl/	http://www.ebi.ac.uk/about/terms-of-use
	CARD	http://card.westgis.ac.cn/	http://card.westgis.ac.cn/about/copyright
	NERC	http://www.nerc.ac.uk/research/sites/data/	http://www.nerc.ac.uk/research/sites/data/policy/
	LTER	http://portal.lternet.edu/	http://www.lternet.edu/policies/data-access
지구환경과학	PANGAEA	http://www.pangaea.de/	https://www.pangaea.de/about/index.php
	AADC	https://data.aad.gov.au/	https://data.aad.gov.au/aadc/about/condition_of_use.cfm
	MGDS	http://www.marine-geo.org/index.php	http://www.marine-geo.org/about/terms_of_use.php
	Seanoë	http://www.seanoë.org	http://www.seanoë.org/html/mentionlegales.htm
물리학	GES DISC	http://disc.gsfc.nasa.gov/	http://disc.gsfc.nasa.gov/additional/ges_disc_data_policy
	NoMaD	http://nomad-repository.eu/	http://nomad-repository.eu/cms/index.php?page=terms-and-conditions
사회과학	Harvard Dataverse	http://dataverse.harvard.edu/	http://dataverse.org/best-practices/harvard-dataverse-policies
	openICSPR	https://www.openicpsr.org/	https://www.openicpsr.org/home/faqs
	UK Data Service - ReShare	http://reshare.ukdataservice.ac.uk/	https://www.ukdataservice.ac.uk/get-data/how-to-access/conditions/eul
일반과학	OSF	https://osf.io/	https://github.com/CenterForOpenScience/centerforopenscience.org/blob/master/TERMS_OF_USE.md
	ADS	http://archaeologydataservice.ac.uk/	http://archaeologydataservice.ac.uk/advice/collectionsPolicy
	Dryad Digital Repository	http://datadryad.org/	http://datadryad.org/pages/policies
	figshare	http://figshare.com/	https://figshare.com/terms

