

디지털 시대 오픈 데이터 정책의 현황과 과제

An Analysis on Trends and Tasks of Open Data Policy in the Digital Era

신은자 (Eun-Ja Shin)*

초 록

과거에는 오픈 데이터에 공감한다 하더라도 이를 실천할 방법이 마땅하지 않았으나 요즘은 디지털 형태의 연구데이터를 IT를 통해 공유하는 것이 어렵지 않은 상황이 되었다. 그러나 많은 연구자가 오픈 데이터를 시행하였을 때의 부작용과 추가 작업에 대한 부담을 느끼고 있고 이외 해결하여야 할 문제도 다소 있어, 오픈 데이터는 현재 기대만큼 활발히 수행되고 있지는 않다. 지구과학, 기상학 등 일부 학문 분야에서 활발하게 추진되고 있을 뿐 나머지 학문 분야에서는 오픈 데이터에 대하여 큰 관심을 보이지 않는 듯하다. 연구결과 해외의 학회, 비영리단체, 대학, 연구지원기관에서는 오픈 데이터를 공공의 이익 추구 차원에서, 주요 출판사에서는 오픈 데이터를 논문을 엄격하게 심사하기 위한 보완책 차원에서 추진하고 있었다. 오픈 데이터는 후속 연구를 이끌고 학문을 발전시키는 발판 역할을 한다는 점에서 중요하고 앞으로 나아가야 할 방향이라는 것은 분명해 보인다. 따라서, 국내에서도 해외의 사례를 충분히 고찰하고 정책에 반영함은 물론이려니와, 연구자, 대학, 도서관 모두 오픈 데이터의 필요성과 향후 전개될 상황에 관하여 관심을 갖고 보다 적극적으로 협력하여야 할 것이며, 이 연구는 이에 관한 구체적인 내용을 기술하였다.

ABSTRACT

There were not many ways to share research data in the past, but modern information technology has allowed us to share these data. As data sharing has its side effects, researchers' attitude and practice to sharing data vary by individual discipline. This study found that foreign learned societies, NGOs, universities and research funders support data sharing in a utilitarian perspective, while major publishers demand it so that other researchers can verify the data in peer review. It is important that open data policy should be settled down in near future for evoking further studies and encouraging progress in science. In order to establish data sharing successfully in Korea, efforts could be made by researchers, universities, academic libraries, and governments as well as the stakeholder. This study also proposed specific ways to perform it.

키워드: 오픈 데이터, 연구데이터, 데이터 아카이빙, 데이터 공유, 리포지터리, 데이터관리계획
open data, research data, data archiving, data sharing, repository, DMP

* 세종대학교 신문방송학과 교수(eunjamail@gmail.com)

■ 논문접수일자: 2015년 7월 23일 ■ 최초심사일자: 2015년 9월 3일 ■ 게재확정일자: 2015년 9월 10일
■ 정보관리학회지, 32(3), 49-68, 2015. [<http://dx.doi.org/10.3743/KOSIM.2015.32.3.049>]

1. 서론

요즘 자신의 연구성과를 학술지에 논문으로 발표하는 연구자가 많다. 우수한 논문을 출판하기 위하여 많은 시간과 노력을 들이고 있고, 연구자간의 논문 출판 경쟁도 점점 더 치열해지고 있는 양상이다. 이에 논문의 우수성을 널리 인정받기 위해서는 실제로 어떤 요건을 충족하여야 하는지를 상기해 볼 필요가 있다. 일반적으로 이들 요건으로는 독창성(originality), 정확성(accuracy), 객관성(objectivity), 불편부당성(impartiality), 독이성(readability), 검증성(verification) 등을 든다(남궁용권, 남궁지영, 2008). 이 가운데 검증성은 후속 연구자가 기존 연구를 재현하고, 문제점을 발견하며, 이를 해결하거나 보완한 연구를 수행함으로써 학문을 발전시킬 수 있다는 면에서 중요하다. 만일 논문에 직접적으로 쓰였던 연구데이터가 잘 보존되어 추가 연구에 재이용된다면, 검증성을 담보할 수 있을 뿐 아니라 이를 바탕으로 후속 연구가 보다 신속하게 수행될 것으로 보인다.

연구자가 많은 시간과 노력을 기울여 수집한 연구데이터가 논문 출판과 더불어 생명을 잃고 사장된다면 학계 전체로 볼 때 적지 않은 손실일 것이다. 1901년 Galton은 연구자간 연구 데이터 공유가 매우 필요하다는 주장을 일찌감치 한 바 있다(Pollard & Celi, 2014). 1932년 'Econometrica'란 학술지 논문의 심사에서도 연구자에게 원 데이터(raw data) 공개를 촉구한 심사의견이 있었다는 기록도 보인다(Dimitrova, 2014). 또한 논문을 읽는 독자 입장에서 볼 때 최종 연구결과물인 논문을 통해

축약된 연구결과만을 보는 것에 만족하지 않고, 원 데이터를 함께 보기 원하는 경우도 부지기수일 것이다. 특히 화학과 같은 분야는 논문과 데이터를 함께 볼 필요가 있고, 기상학과 같이 동일한 조건으로 데이터를 재측정하거나 재수집하는 것이 불가능한 학문 분야는 연구데이터의 공개를 오랫동안 염원해 온 것도 사실이다. 과거에 존재했던 생물을 분석하는 고생물학(palaeontology)과 같이 연구 시간 대부분을 데이터 수집에 할애하여야 하는 학문 분야의 경우에는 연구 데이터의 공개 및 공유에 대한 요구가 상당히 강하다. 이상으로 미루어 볼 때 연구데이터는 조사, 실험, 분석 결과를 확인하고 나아가 재현하고 검증하는데 꼭 필요하며, 장기적으로 학문을 발전시키는 초석이 되는 소중한 자산이라는 데 이견이 없을 것이다.

연구 경쟁의 가속화, 연구 인력의 증가, IT의 눈부신 발전 등으로 인하여 연구 데이터의 양이 기하급수적으로 증가하고 있는 것도 연구데이터를 보존하고 공유하는데 참고하여야 할 내용 중 하나일 것이다. 각종 전자현미경, 실험장비, 인공위성을 통해 매일매일 생산되고 축적되는 디지털 형태의 연구데이터 양 또한 상당할 것이다(김선태, 한선화, 이태영, 김용, 2010). 일례로 유럽원자핵공동연구소(European Organization for Nuclear Research)에서 대용량 소립자 충돌형 가속기란 장비에서 수집하는 데이터만 해도 매년 16 페타바이트(Petabyte, PB)에 달하므로 여러 첨단 장비를 통해 엄청난 양의 데이터가 매년 생산되고 축적되고 있다는 것은 쉽게 짐작할 수 있다. 이처럼 상당 규모의 디지털 형태의 연구데이터가 끊임없이 생산되고 축적되고 있다는 것은 환언하면, 연구데이터가

널리 세계의 연구자에게 공유되고 재이용될 경우 그 효과는 몇 십에서 몇 천 배가 될 것이라는 추정을 가능하게 한다. 과거에는 연구데이터를 공개하거나 출판하여 공유하는 것이 비현실적인 것으로 여겨져 왔다면, 지금은 첨단 기술을 이용하여 연구데이터를 공개하거나 출판하는 것이 순식간에도 가능해진 것이 사실이다.

아울러 연구업적 평가의 객관성을 유지하기 위하여 연구자의 연구업적을 양적으로 평가하는 대학, 연구소, 연구지원기관의 수가 국내의 적으로 증가하고 있다는 점도 연구데이터 공유에 직간접적인 영향을 줄 것으로 보인다. 그동안 연구업적의 양적 평가에 논문의 수, 특허 수 등 비교적 단순한 지표를 이용하는 것이 일반적이었으나, 연구의 수준까지 고려할 수 있다는 점에서 타 연구자에게의 인용빈도를 포함하는 기관이 점차 증가하는 추세이다. 최근에는 연구비 지원을 할 때 연구 중 생산된 연구데이터의 관리와 공유 계획(Data Management Plan, DMP)을 보고 받는 연구지원기관도 생겨나고 있다. 이는 연구지원기관으로 하여금 연구비 예산의 효율적 집행과 지출의 타당성을 확보하게 하는 부수적인 효과도 갖게 해준다.

이와 같은 여러 가지 이유로 인하여 연구데이터의 공개, 공유, 출판은 연구자 뿐 아니라 연구지원기관, 출판사 모두의 관심사가 되었고, 넓게 보아 연구데이터는 이용자에게 제공하여야 할 자료의 일종이기에 이를 적극 수용하는 대학도서관도 생겨나고 있다. 연구데이터의 공개와 출판이 잘 되기 위해서는 해결해야 할 문제가 여럿 있을 뿐 아니라 해당 주체의 상호 이해와 협력이 매우 필요한 게 사실이다. 그런데 이들은 각자가 처한 입장과 이해관계가 달라

서로 협력하는데 있어 자주 한계를 보인다. 현재 연구데이터의 공개와 출판이 지연되거나 안 되는 상황도 자주 나타나고 있고, 결과적으로 연구 데이터의 재이용을 기반으로 한 후속 연구도 활발하지 않다. 그러나 학내의 안팎에서 오픈 데이터에 대한 요구가 점차 커지고 있고, 오픈 데이터가 되지 않는 이유에 대한 비판도 점점 더 거세지고 있는 상황이다. 따라서 이해 당사자들이 문제 해결의 어려움을 들어 계속 소극적인 입장을 취하기는 사실상 어려울 것으로 보인다. 이후 자의든 타의든 이해당사자가 적절하게 타협할 수밖에 없는 상황이 된다면, 연구데이터 아카이빙과 출판도 점차 속도를 낼 것이고, 연구 커뮤니케이션 생태계도 지금과는 확연히 다른 모습을 보일 것이라는 예측도 있다(Murphy, 2014).

이와 같은 시대적인 상황을 배경으로 이 연구는 디지털 시대의 오픈 데이터 정책의 현황과 과제를 살펴보았다. 이 연구에서 관심을 갖고 있는 데이터는 연구데이터로 연구를 수행하는 과정 중에 생산된 다양한 유형의 사실적 기록을 포함한다. 구체적으로 연구데이터는 각종 수치, 수식, 이미지, 도표, 문서 등의 형태를 띤다. 요즈음 연구데이터와 더불어 과학데이터란 용어도 자주 등장하는데, 과학데이터는 관측, 감시, 조사, 실험, 연산, 분석 등 일련의 과학기술 연구 활동의 결과로 얻어진 수치, 공간, 도표, 문서 등을 뜻한다. 결국 연구데이터와 과학데이터는 정의 대부분이 서로 중복되고 두 용어를 세밀하게 구분할 필요성도 낮아(김선태 외, 2010; 김문정, 김성희, 2015), 이 연구에서는 두 용어 의미를 포괄하여 '연구데이터'란 용어로 통칭하였다.

오픈 데이터는 낱알의 데이터가 경제적인 제약 없이 무료로 자유롭게 이용, 재이용, 재배포 되는 것을 의미한다(Royal Society, 2012). 별도의 기술이 없는 한 이 연구에서 '오픈 데이터'는 연구에 직접 이용된 연구데이터를 공유하는 것을 뜻하며, 결과적으로 '오픈 연구데이터'라는 한정된 의미를 갖는다.

이 연구는 일련의 연구 활동으로 생산된 연구데이터를 공개하고 공유하여 연구자간 학술 커뮤니케이션을 보다 원활하게 함으로써 학문의 발전을 도모하는데 초점을 맞추었다. 현재 오픈 데이터에 관한 국내 연구는 연구데이터 공유 실태의 문제점을 지적하고, 이를 해결하기 위한 국가차원의 공유 시스템 개발을 제안하는데 머물러 있을 뿐, 최근 해외의 데이터 공유와 출판 동향을 반영한 종합적인 분석과 제안에는 이르지 못하고 있다. 이에 이 연구는 해외의 비영리 단체나 주요 출판사가 데이터 공유와 출판에 관하여 어떤 정책을 수립하여 실천하고 있는지를 최근 문헌을 중심으로 살펴보았다. 이로써 국내의 연구자, 학회, 연구지원기관, 정부, 대학, 도서관 등 각 주체로 하여금 데이터 공유와 출판이 당면 과제라는 것을 분명하게 인식하게 하고, 이들이 서로 협력하여 오픈 데이터를 적극 추진하는데 도움이 되는 기초 자료를 제공한다는 데 연구의 의의를 두었다.

2. 이론적 고찰

2.1 오픈 데이터의 필요성

현대에는 과학적이고 객관적인 연구를 수행

하는 데 있어 연구데이터의 수집과 분석이 필수로 간주된다. 첨단 연구 장비를 통해 생산되는 연구데이터의 수량도 점점 더 증가하고 있고, 이들 연구데이터는 과거의 그것과 특성 면에서 분명한 차이가 있기도 하다. 이에 이 연구는 본격적인 분석에 앞서 요즘 생산되고 있는 연구데이터의 특성에 대하여 먼저 살펴보았다.

요즘 연구데이터는 인문과학, 사회과학, 자연과학, 공학, 의학에 이르기까지 전 학문 분야에서 생산되고 있다. 실제 연구데이터의 콘텐츠는 수치, 통계, 도표, 사진, 로그파일, 문서 등 매우 다양한 콘텐츠로 구성되어 있다. 아울러 수집되고 있는 연구데이터의 형식도 문서형태의 아날로그 형식에서부터 컴퓨터로 읽고 처리할 수 있는 파일, 이미지 등과 같은 디지털 형식에 이르기까지 상당히 다양하다. 이와 같이 연구데이터의 콘텐츠, 형식, 분량이 다양하다는 것은 보존 및 관리가 결코 용이하지 않음을 시사한다. 아날로그와 디지털 형식이 혼재하고 데이터의 수량도 매우 다양하여 축적하는 것 자체도 단순하지 않은데다가, 추후 재이용까지 염두에 두어야 하므로 보다 과학적이고 체계적으로 수집, 저장, 관리하는 방법이 필요한 것이다.

실제로 현대 학문은 누적된 데이터를 기반으로 한 데이터 분석에 많은 비중을 두고 있다. 따라서 데이터를 공개하고 공유하는 것은 타 연구자에게 직접적인 도움이 되는 것은 물론이려니와 학문과 산업의 발전에도 긍정적인 영향을 줄 것이며 결국 인류 모두에게 보탬이 될 것이다. 구체적으로 오픈 데이터가 주는 잇점을 선행연구를 참조하여 정리하면 다음과 같다(박성욱, 한선화, 2011).

첫째, 데이터를 증장기간 보존하고 재이용

할 수 있다. 특히 반복할 수 없거나 반복에 많은 시간과 비용이 드는 실험 및 관찰의 경우 연구 데이터의 보존과 재이용은 더욱 절실하다. 데이터를 공유하게 되면 여러 브랜드에서 동일한 약을 개발하는데 소요되는 연간 3억 파운드의 경비를 절감할 것이라는 구체적인 보고도 있다 (Pollard & Celi, 2014). 데이터의 보존과 공유가 원활하면 데이터가 더 이상 생산되지 않는 극단적인 상황이 되더라도 추후 이를 고스란히 재이용할 수 있다.

둘째, 보존된 연구데이터를 이용하여 반복 실험을 할 수 있고 선행연구의 검증도 원활해진다. 즉, 오픈 데이터는 선행 연구를 검증하고 문제점을 찾아내며 새로운 방법을 고안하는 차기 연구의 시발점이 될 수 있고, 반대의 경우에는 부정적인 결과가 초래될 것이다. 일례로 미국의 경우 연간 44,000~98,000명의 환자가 불필요한 처치, 오류 반복 등의 과오 처치를 받는 데, 그 원인 중의 하나가 데이터 공유가 미흡한 것이라는 지적이다(Pollard & Celi, 2014).

셋째, 여러 개의 연구데이터를 모아 비교연구를 하거나 여러 데이터를 결합한 메타연구도 시도할 수 있다. 오픈 데이터가 시행되면 이전에 생산된 데이터와 추후 수집한 데이터를 연계해 보다 종합적이고 심층적인 연구를 수행할 수 있을 것이다. 게놈 프로젝트(genome project)와 같은 범세계적인 데이터셋 구축은 종전에 하지 못했던 광범위한 분석도 가능하게 해 준다. 또한 연구자간 협력을 증진시켜 학문 분야 안팎의 공동연구를 촉진할 수 있다. 인접 학문 분야와 다학제적인 연구도 수행할 수 있고 이로 인한 시너지 효과도 기대할 수 있는 것이다.

넷째, 연구데이터의 보존은 교육에도 활용될

수 있다. 데이터를 직접 활용하고 실습수업까지 병행하는 교육을 실시한다면 학습효과는 배가될 것이다. 이 경우 오픈 데이터는 새로운 연구자를 육성하는 자양분의 역할을 하는 것이다.

이상에서 볼 때 오픈 데이터 정책을 장기간 꾸준히 추진하면 기존 연구도 탄력을 받게 되고 보다 나은 학문의 발전도 기대할 수 있다. 즉, 오픈 데이터 정책을 추진하는 것은 연구와 학문을 발전시키는 중요한 발판을 마련하는 것이고, 이것이 오픈 데이터 정책을 꾸준히 추진하여야 하는 주된 이유인 것이다.

2.2 선행연구

현재까지 수행된 오픈 데이터에 관한 외국의 선행연구는 오픈 데이터의 필요성, 오픈 데이터의 효과, 오픈 데이터의 추진 현황, 오픈 데이터의 활성화 정책 제안 등이 주를 이루고 있다. 내용별로 주요 선행연구를 살펴보면 다음과 같다.

오픈 데이터의 필요성과 효과를 기술한 연구로 Kansa와 Kansa의 연구가 있다(2013). 이들은 논문을 통해 연구성과를 전달하는 것보다 원 데이터(raw data) 자체를 직접 공유하는 것은 차원이 다르다고 기술하고 있다. 간단한 예로 데이터를 요약하거나 논문을 작성하면서 생긴 데이터의 누락이나 오기의 경우 원 데이터 공개가 없으면 독자가 오류를 알아내기 어렵다. 반대로 데이터가 공개되어 있다면 추후에라도 이를 바로잡을 가능성이 있다는 것이다. 기존 원 데이터를 이용하여 보다 심층적인 연구를 수행할 수 있다는 것도 오픈 데이터의 긍정적 효과라 할 것이다. 한 걸음 더 나아가 원 데이터를 주기적으로 출판하게 되면 데이터를 계속

수정하고, 설명을 추가하며, 발전시킬 수 있다는 것이다. 이와 같이 학계의 연구자가 데이터를 공유하게 되면 재이용 결과 뿐 아니라, 데이터의 '인용'도 이루어져 데이터를 생산한 연구자에게 경제적, 학술적, 사회적 보상도 주어질 수 있다. 이 이외에 오픈 데이터는 원 데이터를 다른 연구자와 협업으로 갱신하고, 보다 나은 공동연구를 수행할 수 있으며, 이를 계기로 학문의 획기적인 발전도 기대할 수 있다(Kratz & Strasser, 2014).

이와 같이 오픈 데이터는 필요성과 효과는 모두 쉽게 공감할 수 있지만, 현재 오픈 데이터는 학문분야에 따라 매우 다양한 진행속도를 보이고 있다(Kim, 2013). van den Bos 등(van den Bos, Jenny, & Wulff, 2014)은 교신저자에게 연락하여 73%나 데이터 수집에 실패한 적도 있고, 반대로 85% 넘게 제공받은 예도 있다고 보고한 바 있다. 이와 같이 데이터 수집에서 큰 차이가 난 것은 요청 받은 교신저자 개인이 오픈 데이터에 얼마나 협조적인가와 관련이 있지만, 학문 분야의 특성상 데이터 공유가 얼마나 절실한가와도 밀접한 관계가 있을 것이다. 다시 재현하거나 검증하는 것이 거의 불가능한 학문 분야에서는 데이터 요청을 거절하기 어렵고, 더불어 자신도 언제인가 타 연구자에게 데이터를 요구할 상황이 올 수도 있다는 생각을 하기 쉽다.

지구과학에서는 동일 조건의 관측이나 복제가 사실상 불가능하다(Callaghan, 2014). 지구과학 데이터의 경우 관측기간도 각각 다르고 분량도 매우 다양하다는 특징을 갖고 있다. 이에 비해 기상학은 이론적으로 가능하나 실제 상황처럼 엄격하게 재추정하는 것이 불가능에

가까울 수 있다. 이러한 상황을 고려할 때 이 두 분야에서의 데이터 공유에 대한 요구와 기대는 다른 분야에 비해 상대적으로 크다. 따라서 이들 분야에서 연구데이터를 공유하는 것은 일상적이고 자연스러운 일로 보아야 할 것이다.

고생물학(palaeontology)도 오픈 데이터에 관한 기대가 큰 학문 분야 중의 하나이다(Mounce, 2014). 고생물학은 과거에 존재했던 생물의 껍적을 연구하므로 데이터의 수집이 전체 연구에서 차지하는 비중이 매우 높다. 최근 들어 첨단 IT를 통해 고생물에 관한 다양한 디지털 데이터를 수집하게 된 것은 이 분야 연구자에게 큰 희망을 주고 있다. 각각의 조각에 불과했던 데이터도 때로는 고생물의 생태 흔적을 밝히는 중요한 단서가 되기도 하고 결과적으로 조사, 분석, 탐구를 촉진하는 도화선이 되기도 한다. 이 분야의 유명 학술지인 '*Journal of Paleontology*'와 '*Paleobiology*'는 논문의 저자에게 Dryad라는 리포지터리(repository)에 데이터 아카이빙을 하도록 요구하는 등 이 분야의 오픈 데이터 정책은 매우 구체적이고도 현실적이다.

경제학 분야도 데이터가 갖는 중요도가 커서 데이터 공유에 관한 연구자의 요구가 많다. Dimitrova(2014)는 소수이지만 연구자가 자신이 수집한 데이터를 웹사이트에 올려 후속 연구의 출발점이 되게 하는 경우도 있다고 보고한 바 있다. 경제학 학술지 '*American Economic Review*'는 경험적 논문의 경우 원 데이터와 초기 분석결과를 저자로부터 받아 웹사이트에 꾸준히 등재하고 있다(Kim, 2013). 그러나 대체로 이 분야 연구자는 데이터 공유에 관하여 적극적인 태도를 취하고 있지는 않는 듯하다. 그 이유로 이들은 개인 프라이버시, 국가 기밀(보

안), 상업적 보안 등을 들기도 한다. 데이터를 다른 연구자와 공유하려면 생산된 형태 그대로가 아니라 코드별로 설명을 달아 재이용이 원활하도록 해야 하는데, 여기에 소요되는 시간과 노력이 만만치 않다는 것이다. 반면에 데이터 공유시 직접적으로 개인 연구자에게 주어지는 보상과 혜택도 분명하지 않아 적절한 동기 부여가 되지 않는다는 것이다.

심리학 분야에서는 미국심리학회(American Psychological Association)에서 데이터를 공유를 권하고 있다는 점이 주목을 끈다(van den Bos, Jenny, & Wulff, 2014). 그러나 아직 데이터 공유를 위한 데이터 포맷 설정, 라이선싱, 아카이빙 방법 등을 구체적으로 명시하지는 않았다는 점에서 아쉽다고 할 것이다. 미국심리학회는 오픈 데이터의 엠바고를 5년으로 정하고 있는데, 이들은 데이터 공유의 실효성을 살리기 위해서는 이 기간을 좀 더 단축하여야 할 것이라고 지적하였다. 이 이외에 데이터를 오픈한 논문이 인용에 유리하다는 것을 알고는 있지만, 데이터 공개를 위해서는 별도의 설명을 추가하여야 하므로 연구자에게 보상을 추가하여야 한다는 의견이 지배적이다.

학문분야에 따라 오픈 데이터에 관한 관심이 다르긴 하지만 오픈 데이터 아카이빙의 추진 효과가 어느 정도인지에는 공통적으로 관심이 쏠리고 있다. Vines 등(2013)은 데이터 아카이빙 정책의 강도에 따라 학술지를 네 집단으로 나누어 그 효과를 분석하였다. 분석 결과 아카이빙 정책이 있는 학술지는 그렇지 않은 학술지에 비해 데이터 입수 가능성이 25배까지 더 높았다. 데이터의 입수 가능성은 아카이빙 정책이 없는 경우에 비해 비명문화 의무적 아

카이빙 정책은 17배, 명문화 의무적 아카이빙 정책은 974배 더 높았다. 한편 아카이빙 권고 집단의 아카이빙 비율은 아카이빙 정책이 없는 경우에 비해 약간 높을 뿐이었다. 결과적으로 아카이빙을 논문 투고 요령에 명문화하는 것은 저자에게 데이터 아카이빙을 적극 참여하게 하며 오픈 데이터 성공에 매우 효과적이라는 것을 확인할 수 있다.

국내의 오픈 데이터에 관한 선행연구는 해외 오픈 데이터 사례 소개, 국내 연구데이터의 관리 현황 분석, 국내 연구데이터 공유 시스템의 개발 제안에 집중되고 있다.

윤종민과 김규빈(2013)은 주요국 연구데이터의 공개 및 공유에 관한 법률 규정을 어떻게 마련하고 있는지를 조사한 바 있다. 연구데이터의 오픈 액세스 실천을 위해 미국은 경쟁력 강화법(America Competes Reauthorization Act), 호주는 책임연구수행법(Australian Code for the Responsible Conduct of Research)을 시행 중에 있고, 중국은 과학데이터제출입시법과 과학데이터공유조례를 입법 중에 있다. 이들 세 나라는 국가차원에서 연구데이터 관리 중장기 계획을 수립하고, 단계적으로 연구데이터관리사업을 추진하고 있으며, 법제도를 계속 강화해 나가고 있는 것으로 확인되었다.

국내 연구데이터의 공유는 매우 제한적으로 시행되고 있다는 보고도 있다(이상환, 심원식, 2009; 김지현, 2012). 한국과학기술정보연구원(2012)이 한국에너지기술연구원을 비롯한 9 곳의 데이터 담당자와 인터뷰를 실시한 결과, 연구데이터는 연구자 개인이나 연구팀 단위로 축적하고 관리하고 있으나, 공유되거나 재이용되는 경우는 극히 드물었다. 한 곳을 제외하고는

데이터 전달 인력이나 부서가 존재하지 않았고, 소정의 메타데이터도 작성하고 있지 않았다. 신영란과 정연경(2012)은 기초학문자료센터(Korea Research Memory)를 비롯한 국내외 9곳의 아카이브 현황을 조사하고 면담을 수행하였다. 그 결과 KRM은 연단위로 사업단을 선정해 사업의 연속성이 약한 것이 심각한 문제였고, 재이용할 만한 연구데이터도 많이 부족하였다.

기관별 데이터 아카이브에 관한 보고도 소수 있었다. 신영란(2011)은 한국사회과학도서관의 소장장서와 데이터 아카이브를 함께 서비스하는 한국사회과학자료원(Korea Social Science Data Archive, KOSSDA)의 구축 과정, 서비스 내용, 향후 계획을 밝힌 바 있다. 이에 의하면 KOSSDA는 기존 도서관에서 수행하던 문헌 서비스 뿐 아니라 연구 데이터까지 제공하는 차원 높은 서비스를 연구자에게 제공하기 시작하였다. 그러나 본 연구에서 최근 상황을 확인한 결과 현재 양적 데이터는 1,900건, 기록 데이터는 200건 정도여서, 축적된 데이터 양이 많지 않다는 한계가 있었다(KOSSDA, 2015). 성진석(2011)은 연구데이터를 기관 차원에서 수집하고 아카이빙하기 위하여 한국농촌경제연구원 연구관리규정을 2011년 개정하였다고 보고하였다. 이로써 그는 연구자 개인이 관리하던 연구데이터를 기관에서 중앙집중적으로 수집하고 보존할 수 있는 법적 근거가 마련되었고, 앞으로 전자도서관시스템을 통해 연구데이터를 제공할 계획이라고 밝혔다.

강희중(2012)은 국내 연구데이터 생산에 투입되는 예산에 비해 활용이 낮은 것은 국가차원의 전략이 미흡한데 그 원인이 있다고 분석하였다. 이와 같은 상황을 극복하기 위해서 과

학기술정보연구원(2012)은 학문 분야별 리포지토리를 구축하기 보다는 여러 연구 분야에서 활용 가능한 범용 리포지토리를 개발하는 것이 바람직하다고 하였다. 이에 앞서 박성욱과 한선화(2011)는 전자 대신 후자와 같은 방식을 취할 경우 생산유발 효과가 1.5배 높고, 부가가치 유발과 취업자 유발 효과도 각각 전체 산업 평균을 상회할 것이라는 분석결과를 발표하기도 하였다.

이상에서 볼 때 학문 분야에 따라 차이가 있기는 하지만 외국의 경우 데이터 아카이빙과 공유에 관한 정책을 이미 오래 전 수립하고 실천하면서 연구자의 참여를 점점 더 확대해 나가고 있는 반면, 국내 연구는 연구데이터 공유 실태의 문제점을 지적하고, 이를 해결하기 위하여 기관 또는 국가차원의 공유 시스템 개발에 착수하고 있는 단계에 머물고 있다는 것을 알 수 있다.

3. 유형별 오픈 데이터 추진 정책

과거에 비해 데이터를 공개하고 공유하는 데 대한 인식이 전반적으로 좋아지고 있는 것은 사실이지만, 오픈 데이터는 추진 주체에 따라 속도와 방향에서 차이가 크다. 이에 이 연구는 선행연구를 토대로 비영리단체와 출판사로 나누어 해외의 오픈 데이터 추진 정책 현황을 분석하였다. 이 연구에서는 비영리단체의 범위를 광범위하게 설정하여 수익을 최우선시 하지 않는 모든 단체, 기구, 기관을 포함하였다. 결과적으로 연구자간의 학술교류 증진을 추구하는 전

통적인 학회, 비영리적 목적으로 설립된 단체나 기구, 공공의 이익에 관심을 갖고 있는 정부 부처나 기관, 대학과 대학도서관 모두를 이에 포함하였다. 이 가운데 이 연구는 국내의 오픈 데이터 정책 수립에 직접적인 도움이 될 만한 최근 해외 사례를 중심으로 분석하였다.

3.1 비영리단체의 오픈 데이터 정책

〈표 1〉은 비영리단체의 오픈 데이터에 관한 주요 내용을 보여준다. 전통적으로 학회는 투고된 논문을 엄선함으로써 학술지가 일정 수준 이상의 질을 유지할 수 있게 하는 역할을 한다. 최근에는 투고된 논문의 심사 외에 해당 논문의 원 데이터를 온라인으로 제공하여 후속 연구를 활성화시킴으로써 학회의 기능과 역할을 한 층 높이기도 한다.

일례로 영국생태학회(British Ecological Society)는 데이터 공개를 보다 확실하게 실천하기 위하여 데이터의 아카이빙을 의무화하고 있다. 이 학회는 데이터 아카이빙의 의무화에 관하여 2012년 사전 홍보, 2013년 시범 운영, 2014년 본격 시행을 추진하였다(Norman, 2014). 대부분의 연구자가 연구 중 생산된 데이터도 함께

제출하고 이를 다른 연구자와 공유하는 것이 학문의 발전에 도움이 된다는 것을 알고 있지만, 의무화라는 강도 높은 정책으로 데이터 아카이빙을 추진하는 것에는 거부감을 갖기 쉽다. 자칫 기존의 출판 프로세스를 복잡하게 하고 데이터 공개에 따른 부작용도 불러올 수 있다고 보기 때문이다. 정치적으로 민감한 내용이나 팽팽하게 논쟁 중인 사안의 경우 데이터가 공개됨으로써 분쟁을 초래할 것이라는 염려도 있다. 그러나 생태학의 경우 축적된 과거 데이터에 새로운 데이터를 결합할 경우 의미 있는 연구 성과가 나올 가능성이 높다는 데 연구자간 공감대가 이미 충분이 있어 온 듯하다. 이를 배경으로 영국생태학회는 의무화 정책을 적극 시행하였고, 시행 후 6개월이 지난 지금 많은 연구자가 데이터 의무화 정책을 잘 따르고 있는 것으로 보인다. 의무화 정책 시행 전 보다 투고율도 6.7% 상승한 것으로 집계되어 이 정책은 안정화 단계에 접어든 것으로 확인되고 있기 때문이다.

영국생태학회가 데이터 아카이빙 의무화에 착수하여 비교적 단시간 내에 제자리를 잡은 데는 몇 가지 성공요인이 있다(Norman, 2014). 본격 시행에 앞서 충분히 홍보기간을 두어 연

〈표 1〉 주요 비영리단체의 오픈 데이터 상황

| 구분 | 정책·제도 | | 예산 지원 | | 교육 |
|-------|---------------------------|--------------------|----------------------------|-----------------------------|-------------------|
| | 영국생태학회 | MSF | RCUK | Office of S&T Policy | MIT 대학 |
| 시기 | 2014 | 2012 | 2013 | 2013 | 2009 |
| 주요 정책 | 데이터 아카이빙 의무화 리포지터리 자율화 | 리포지터리에 데이터 기탁 | 공공재원 투자 데이터는 공공재로 광범위하게 공유 | 논문과 데이터 OA에 연간 1억 달러의 예산 지원 | 데이터 관리 지침 제정 및 실천 |
| 특이 사항 | 시행 이후 투고율 6.7% 상승 | 데이터 공유 비용을 적정수준 유지 | | | 데이터 관리 교육 실시 |

구자의 동의를 최대한 이끌어 냈고, 독점 아카이빙을 하지 않고 분산하여 연구자에게 부담을 주지 않았다. 엠바고 기간도 탄력성 있게 하여 연구자의 불만을 최소화 한 것도 주효했던 것으로 보인다. 부연하면 영국생태학회는 연구자가 원하는 곳에 데이터를 아카이빙 할 수 있게 허용하였고, 중복하여 아카이빙 하는 경우도 문제 삼지 않았다. 그 결과 아카이빙된 데이터의 42%는 영국생태학회가 후원하는 Dryad, 28%는 Wiley Online Library, 24%는 공공 리포지토리를 포함한 기타에 아카이빙 되고 있는 것으로 관찰되었다. 실제로 영국생태학회는 2013년 아카이빙 비용으로 Dryad에 £5,500(약 660만원)을 지불한 바 있다. 연구자는 본인이 제출한 데이터를 어느 시점부터 공개할 것인지에 민감하기 마련인데, 영국생태학회는 연구자가 원할 경우 1년까지 엠바고를 둘 수 있게 하였다. 정치적 또는 논쟁중인 문제의 데이터라서 이 보다 더 길게 엠바고를 두기 원하는 경우에는 이를 허용하는 융통성도 발휘하였다.

인도적인 목적으로 설립된 비영리단체에서 오픈 데이터에 관심을 갖는 것은 자연스러운 일일 것이다. 약물의 내성 증가와 유행성 질병의 확산과 같은 사태가 빈번한 요즈음 의료분야에서 데이터 공유가 제대로 되지 않는다면 매우 심각한 문제를 불러 올 수 있다(Karunakara, 2014). 때에 따라서는 국경을 초월한 신속한 공동 대응이 필요하고, 이에 데이터 공유는 더 이상 선택이 아닌 필수로 보아야 할 것이다. 비영리단체인 MSF(Medecins Sans Frontieres, 국경없는 의사회)는 2012년 이후 데이터 공유 정책 하에 정기적인 임상 및 연구데이터를 수집하고 있다. 이들의 오픈 데이터 정책은 데이터

수집이 위협하지 않아야 하고, 정치적인 목적에 쓰이지 않아야 하며, 더 많은 사람에게 건강 혜택을 주는 범사회적인 유용성이 있어야 하고, 경제적 제약이나 지적소유권으로 인하여 데이터 접근이 어렵지 않아야 내용을 담고 있다. MSF는 심사제 논문의 저자, 의료업계 종사자도 데이터 공개를 위해 공공 리포지토리에 기탁하게 함으로써, 오픈 데이터 운영 경험을 계속하여 쌓아가고 있다. MSF는 독지가의 후원금으로 운영되는 만큼 데이터 공유에 들어가는 비용은 과도하지 않은 적정 수준을 유지하도록 하고 있다.

연구자에게 연구비를 지원하는 연구지원기관은 일찍이 오픈 데이터 정책에 관심을 갖고 이를 시행 중에 있다. 캐나다, 영국, 미국 등 주요 선진국도 공공재원으로 수행된 연구결과물은 일반인 모두 쉽게 접근할 수 있는 공공재로 간주하여야 할 것이라는 OECD의 정책을 적극 수용하고 있다. 2010년부터 2012년에 걸쳐 EU는 데이터 공유, 재이용, 보존을 위한 ODE(Opportunities for Data Exchange)를 추진한 바 있다(Royal Society, 2012).

2008년 캐나다는 공적 기금으로 조성된 연구비의 지원을 받아 생산된 데이터는 오픈 액세스로 공개할 것을 의무화 하였다. 연구비 신청시 DMP를 제출하게 하고, 대학을 비롯한 연구비 감독기관에서 이를 모니터링하며, 미준수시 추후 연구과제 선정에 불이익을 주는 등 오픈 데이터 정책을 매우 적극적으로 실천하고 있다(Brown et al., 2015).

영국의 RCUK는 공공재원으로 생성된 연구데이터는 공공재이고 공개를 원칙으로 한다고 기술하고 있다(RCUK, 2013). 이 원칙은 연구

데이터를 광범위하게 공유함으로써 학문적, 경제적, 사회적인 이익을 함께 누리는 것을 포함하고 있다. 영국 Wellcome Trust는 연구데이터는 귀중한 자원으로 보고, 연구가 수행될 때마다 DMP를 제출하도록 하였다.

미국의 NIH는 2003년부터 데이터 공유 정책을 수립하여 실천 중에 있다. 이들은 연구성과를 지식, 제품, 절차로 바꾸기 위해서는 이전이 필수이고 이를 위해서는 데이터 공유가 매우 중요하다고 보았다(Caulfield, Harmon, & Joly, 2012). 연구비 수혜자는 데이터 공유를 위해 연구 기간 언제든지 직접비에 \$500,000 또는 그 이상의 비용을 지출할 수 있도록 제도를 마련하였고, 또한 공유가 불가능할 때는 그 이유를 명기하도록 하였다. 2013년 백악관 과학기술정책부(Office of Science and Technology Policy)는 연방정부기금의 지원을 받은 연구의 경우 논문과 디지털 데이터를 오픈액세스 하는데 연간 1억 달러의 예산을 소요하겠다는 방침을 정해 실행 중에 있다(The White House OSTP, 2013). Bill & Melinda Gates Foundation은 Global Health Data Access Principles에서 지원자는 DMP를 반드시 포함하게 하였고, 가능하면 공공 리포지터리에 아카이빙할 것을 요구하였다. 이것이 불가능하다면 데이터셋 파악이 용이하고, 장기간 보존하는 대안을 제시하도록 하고 있다(Caulfield, Harmon, & Joly, 2012).

선진국의 대학과 대학도서관도 오픈 데이터 정책을 적극 수용하여 시대의 변화에 부응하고 있는 것으로 보인다. 미국과 영국의 주요 대학인 MIT 대학, Wisconsin-Madison 대학, Edinburgh 대학, Southampton 대학 도서관에서는 각 구성원에게 데이터 관리 지침을 줄 뿐 아니라, 연

구자 스스로 데이터를 축적하고 공개할 수 있도록 웹사이트에 셀프 가이드를 제공하고 있다(Kim, 2013). 일례로 이 연구에서 MIT 대학 도서관의 웹사이트를 살펴 본 결과, 이 곳에서는 데이터를 축적하고 공유하는 방법을 쉽고 자세하게 안내하고 데이터 공유의 효과를 적극 홍보하고 있었다(<http://libraries.mit.edu/data-management/share/find-repository/>). 구체적으로 이들은 DSpace, DVN(Harvard Dataverse Network), ICPSR(Inter-university Consortium for Political and Social Research) 등에 데이터를 등재하라고 데이터 생산자인 연구자에게 상세히 알려 주고 있었고, 실제로 데이터 이용자에게는 어떻게 데이터를 찾고 이용할 것인지에 관하여 자세히 설명하고 있었다. 아울러 Minnesota 대학, California 디지털 도서관도 정기적으로 데이터 관리 워크숍을 열어 오픈 데이터를 홍보 및 교육하고 있었고 비공식적인 대민 서비스도 자주 하는 것으로 보고되고 있다(Kim, 2013).

이상에서 볼 때 해외의 경우 생태학과 같이 오픈 데이터에 관한 공감대가 뚜렷한 분야에서 학술지 논문 투고시 데이터도 함께 제출하게 하는 데이터 아카이빙 의무화가 이미 추진되고 있음을 알 수 있다. 통상적으로 주요 선진국에서는 연구 초기 단계부터 데이터 아카이빙에 소요되는 비용을 DMP에 반영하여 연구지원기관이나 정부에서 적극 지원하고 있는 것도 확인할 수 있다. 데이터의 관리, 축적, 공유에 관한 방법 등 이와 데이터와 관련된 전반적인 교육은 대학도서관에서 주도적으로 실시하고 있다는 것도 알 수 있다.

3.2 주요 출판사의 오픈 데이터 정책

주요 출판사의 오픈 데이터 정책의 사례는 <표 2>와 같다. 과학·공학·의학 협회(STM Association)는 2007년 브뤼셀 선언을 통해 논문 출판에 이용된 원 데이터는 모든 연구자에게 무료로 제공되어 검증과 재이용에 활용되어야 한다는 입장을 밝혔다(Murphy, 2014). 회원 출판사는 협회의 취지를 이해하고 따르기 바라며, 저자가 논문을 투고할 때 데이터 세트도 함께 제출하도록 한다면 실현 가능성은 충분하다고 하였다. 과학·공학·의학 협회는 2012년 DataCite와 공동으로 메타데이터의 표준을 선정함으로써 데이터 출판의 틀을 공고히 하였고, 출판사에서 데이터 출판을 다룰 것을 요구하였다. 이들은 논문의 저자가 단순히 논문을 출판하는 것으로 제 역할을 다 하였다고 볼 것이 아니라, 연구 중 생산한 데이터를 공신력 있는 아카이브에 기탁하여 후속 연구를 이끌어야 한다는 점을 분명히 하였다.

실제로 2013년 이후 유명 출판 단체에서 논문 투고시의 데이터 제출 지침을 마련해 시행하기 시작하면서 데이터 출판도 제 속도를 내기 시작하였다. Nature Publishing Group과 PLOS(Public Library of Science, PLOS)가

대표적인 사례이다. Nature Publishing Group (2013)의 'Nature' 학술지는 여러 세부 분야를 다루고 있지만, 특히 생명과학의 경우 오픈 데이터를 다음과 같이 구체적인 절차를 통해 실천하고 있다.

생명과학 논문은 실험 분석 설계를 논문에 상세히 밝혀야 한다고 투고 지침을 통해 분명하게 밝히고 있다. 이는 연구결과를 투명하게 보고하고 재현성을 확보하지는 차원이라고 설명하고 있다. 아울러 논문심사가 이루어지는 동안 편집위원장과 심사위원이 데이터를 직접 확인할 수 있게 저자가 적극 협조할 것으로 요구하고 있다. 이들은 직접 데이터를 제출받는 방식 대신 저자로 하여금 데이터를 리포지터리에 등재하도록 주문하고 있다. 이 경우 심사위원은 익명으로 접근할 수 있지만 기밀 유지에는 각별히 유념하라는 지시를 준수하여야 한다.

<표 3>은 'Nature' 학술지가 정한 데이터 리포지터리로, 이는 논문의 세부 주제에 따라 리포지터리가 각각 다르다는 것을 보여주고 있다. 그러나 'Nature' 학술지는 상황이 여의치 않다면 Dryad나 Figshare와 같은 범용 리포지터리에의 등재도 허용하고 있다. 투고된 논문이 심사를 통과하고 출판일까지 정해지면 저자는 리포지터리와 연락하여 외부로 연구데이터를 공

<표 2> 주요 출판사의 오픈 데이터 상황

| 구분 | STM Association | Nature | PLOS |
|-------|-------------------------|----------------------------|--|
| 시기 | 2007 | 2013 | 2014 |
| 주요 내용 | 논문투고시 데이터세트 제출 | 생명과학 논문 투고시 데이터를 리포지터리에 등재 | 투고시 데이터제공 동의 서약 리포지터리에 데이터 등재 권장 데이터세트 접근번호 제출 |
| 의의 | 메타데이터 표준 선정 데이터 출판 틀 완성 | 심사위원의 데이터 참조 가능 | 독자의 데이터 접근 지원 |

〈표 3〉 'Nature' 학술지의 데이터 아카이빙 리포지터리 현황

| 분야 | 리포지터리 명칭 |
|---|--|
| Protein sequences | Uniprot |
| DNA and RNA sequences | Genbank DNA DataBank of Japan(DDBJ) EMBL Nucleotide Sequence Database(ENA) |
| DNA and RNA sequencing data | NCBI Trace Archive NCBI Sequence Read Archive(SRA) |
| Genetic polymorphisms | dbSNP dbVar European Variation Archive(EVA) |
| Linked genotype & phenotype data | dbGAP The European Genone-phenome Archive(EGA) |
| Macromolecular structure | Worldwide Protein Data Bank(wwPDB) Biological Magnetic Resonance Data Bank(BMRB) Electron Microscopy Data Bank(BMDB) |
| Microarray data | Gene Expression Omnibus(GEO) ArrayExpress |
| Crystallographic data for small molecules | Cambridge Structural Database |

개하여야 하는 책무를 갖게 된다. 'Nature' 학술지는 저자로 하여금 데이터 디스크립터를 'Scientific Data'에 출판하게 함으로써 연구결과의 투명성을 높이고 데이터를 재이용하는데 앞장서고 있다. 이들은 논문 출판 전후에 데이터 디스크립터를 출판하는 것이 논문의 보완 기능을 하므로 꼭 필요하다는 입장이다. 그럼에도 불구하고 리포지터리에 연구데이터를 등재하지 못하는 불가피한 경우 저자는 논문 투고시 추록파일(supplementary information file)을 등재하는 방법을 택할 수 있다. 이 경우에도 논문 출판일이 확정되면 저자는 연구데이터 요구시 URL을 제공하여야 하는 책무가 있다.

PLOS(2013)는 2014년 이후 학술지에 논문을 출판하는 저자는 연구성과를 뒷받침하는 데이터를 조건 없이 제공하여야 한다는 내용의 데이터 정책을 실행에 옮기고 있다. 논문을 투고할 때 저자로부터 데이터 제공에 동의한다는

내용의 서약을 받고 있고, 데이터 공유에 불응하는 것은 게재불가의 근거가 된다고 밝히고 있다. PLOS는 저자로 하여금 데이터 접근에 적극 협조하도록 하고 있고, PLOS에 출판된 논문의 데이터에 접근하는데 어려움이 있는 독자는 편집위원회에 알리도록 하여 저자와 독자 사이에서 데이터 접근의 교량 역할을 하고 있다. 한 걸음 더 나아가 이들은 끝내 데이터 제공에 협조하지 않는 경우 PLOS는 논문 게재 취소도 할 수 있다고 입장을 분명히 밝히고 있다. PLOS가 권장하는 데이터 공개 방법은 적합한 공공 리포지터리에 등재하는 것이다. PLOS 학술지의 저자는 리포지터리가 다루고 있는 주제와 데이터 구조를 참조하여 적합한 리포지터리를 선택하여야 한다. 마이크로어레이(microarray)는 ArrayExpress 또는 GEO, 유전자염기서열은 GenBank, EMBL, DDBJ, 생태학은 Dryad 등에 등재할 것을 권하고 있다. 기타 Dryad와 같

은 범용 리포지토리를 이용하는 것도 허용한다는 것이 PLOS의 입장이다. 저자는 투고시나 늦어도 논문심사 통과시에는 리포지터리 명칭, DOI, 데이터 세트 접근번호 등을 제공하여야 하고, 소규모의 데이터 셋인 경우 논문과 더불어 추록 파일을 제출하도록 하고 있다.

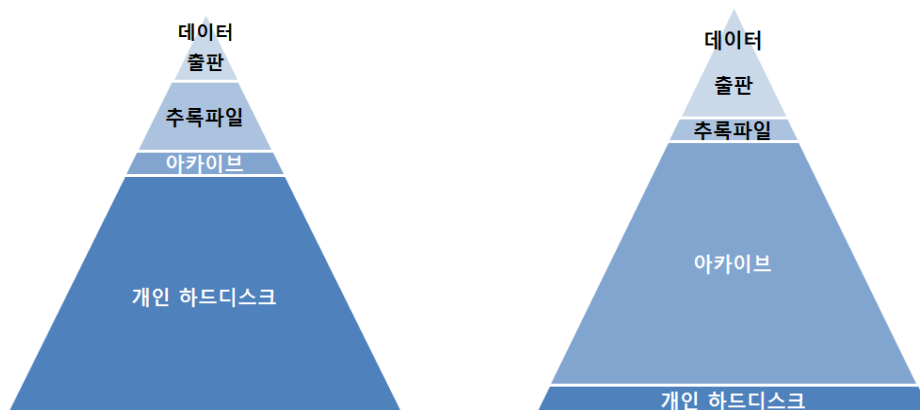
아울러 데이터 디스크립터를 수록한 정기간행물도 속속 등장하고 있다. 'Nature'의 'Scientific Data' 외에도 Wiley사의 'Geoscience Data Journal', 'F1000 Research Journal', BMC의 'GigaScience' 등이 이들 정기간행물에 해당된다. 이상에서 볼 때 최근 주요 출판사는 데이터 공개에 많은 관심을 보이고 있고 오픈 데이터 정책을 적극 실천하기 시작한 것으로 보인다. 이렇게 되면 데이터 공개의 부작용을 우려하여 소극적이었던 여타의 출판사도 데이터 출판을 점차 당면과제로 받아들일 것으로 예상된다.

3.3 논의

분석결과 논문의 오픈 액세스는 과거에 비해

활발한 데 비해, 데이터를 자유롭게 이용하는 데이터의 오픈 액세스는 시작 단계를 많이 벗어나지 못한 것으로 보인다. <그림 1>은 오픈 데이터 정책 시행 전후의 오픈 데이터 현황을 비교한 것으로 이를 통해 미래의 오픈 데이터 모습을 짐작해 볼 수 있다(Murphy, 2014). 그림 왼편은 오픈 데이터가 활발하지 않은 경우로 연구데이터의 상당량이 연구자의 하드 디스크에 저장되어 있음을 알 수 있다. 그림 오른편은 많은 연구자가 오픈 데이터 정책에 참여한 경우로 연구자의 하드 디스크에 저장되어 있는 데이터는 감소하고 데이터 아카이브가 대폭 증가하였음을 보여준다.

김지현(2013)의 분석에 의하면 현재 해외의 연구비지원기관 15곳 중 11곳이 DMP를 의무적으로 제출하게 하고, 생산된 지 6~12개월 이내에 데이터를 공개하며, 데이터 보존 기간은 3년, 5년, 10년 등 다양한 상황임을 알 수 있다. 이처럼 해외의 경우 선도적인 연구비지원기관을 중심으로 실천 가능한 오픈 데이터 정책을 이미 시행하고 있다. 그렇지만 해외라



<그림 1> 오픈 데이터 정책 시행 전후의 데이터 보존 현황

하더라도 오픈 데이터에 관한 인식이 학문 분야마다 차이가 많고 데이터 공유에 필요한 사전 작업에 소요되는 노력과 시간도 만만치 않아 이에 대한 적절한 보상이 주어지지 않는 한 오픈 데이터가 빠르게 확산되기 어렵다고 보는 견해가 일반적이다. 국내의 오픈 데이터는 이보다 더 미진한 상황이고 향후 해결하여야 할 문제도 적지 않다. 이에 국내 오픈 데이터의 당면 문제에 대한 해결 및 추진 방안을 정리하면 다음과 같다.

첫째, 연구자가 오픈 데이터를 잘 이해하고 적극 참여할 수 있도록 분위기를 조성하고 제도를 보완하여야 한다. 현재 많은 연구자가 오픈 데이터에 소극적으로 참여하고 있는데, 이의 원인은 크게 두 가지로 보아야 할 것이다. 우선 연구자가 오픈 데이터를 하였을 때 부작용에 대한 우려를 들 수 있다. 혹시 발견될지 모르는 데이터 오류에 대한 지적이 부담스러울 수 있고, 다른 연구자가 자신의 데이터를 이용해 재빨리 후속 연구를 수행함으로써 연구의 주도권을 빼길 수 있다는 불안감도 가질 수 있다. 다른 하나는 연구자가 데이터를 공개할 때 필요로 하는 추가 작업이 부담스럽다는 것이다. 오픈 데이터는 다른 연구자의 재이용을 염두에 두는 것이므로 언제, 왜, 어떤 방식으로 데이터를 수집하고 축적하였는지에 대한 별도의 설명을 추가하는 등 적지 않은 시간과 노력을 요하기 때문이다.

이와 같은 상황이지만 만일 오픈 데이터가 가져다주는 혜택이 연구자의 우려나 부담보다 크다면 연구자의 오픈 데이터에 관한 인식은 점차 바뀔 수도 있다. 예를 들어 연구자가 논문을 출판해 연구업적을 인정받는 것처럼 데이터

공개도 연구업적으로 인정되는 등의 보상이 주어진다면 연구자의 태도는 달라질 수 있을 것이다. Callaghan(2014)은 영국의 한 학술회의 발표에서 응답자의 67%는 데이터 공개를 통해 연구업적을 인정받을 수 있다면 기꺼이 데이터를 리포지터리에 기탁하겠다는 분석결과를 발표하였다는 것을 상기할 필요가 있다. 현재 국내 대학을 비롯한 많은 기관에서 논문의 수 뿐만 아니라 논문의 피인용을 연구업적의 평가요소로 삼고 있는데, 앞으로 데이터의 공개 및 인용도 이에 포함시켜 가산점을 주는 등의 인센티브가 주어진다면 분위기는 점차 개선될 것으로 보인다.

둘째, 연구지원기관은 연구비를 지원하면서 데이터 공개를 의무화하는 정책을 수립하여 구체적으로 실천하여야 한다. 외국에서는 공적 재원이 투입된 연구의 경우 논문의 오픈 액세스와 더불어 데이터 공개를 요구하는 곳이 계속 증가하고 있고 연구자들은 이미 이것을 자연스러운 일로 받아들이고 있는 분위기이다. 외국은 이미 연구비 신청 단계부터 DMP를 요구하고 오픈 데이터에 참여하도록 한 연구지원기관이 꾸준히 증가하고 있어, 오픈 데이터에 참여하는 연구자 수는 계속 증가할 것으로 보인다. 데이터 수집 및 관리에 만전을 기하기 위하여 공공 리포지터리에 데이터를 아카이빙하게 하는 경우가 많은데, 연구지원기관은 이에 대한 적절한 재정 지원을 하여야 할 것이다(Murphy, 2014). 범용 리포지터리인 Dryad의 연간 운영비가 40만 달러에 이르고 추가 및 큐레이션에 1만 달러가 별도로 소요될 정도로 비용 부담이 높아 연구지원기관이나 대학의 지원 없이 연구자 개인이 이에 데이터를 아카이빙 하여 오픈

데이터에 직접 참여하는 것을 기대하는 것은 현실적이지 않다. DMP 의무화와 더불어 연구 지원기관은 데이터 관리, 축적, 공유에 예산지원을 보다 확실하게 하여야 할 것이다.

셋째, 대학과 대학도서관은 구성원이 데이터를 관리하고, 축적하고, 공유하는데 실질적인 도움을 주어야 할 것이다. 이에 관한 지침을 정하고 정기적으로 교육을 실시하여 이를 구성원에게 주지시키며, 축적된 데이터를 학내외에 적극 알리는 역할을 수행하여야 할 것이다. 이로써 대학도서관은 인쇄 자료, 전자 원문에 이어, 디지털 데이터까지 장서로 인식하고 서비스를 확대하여 시대에 부응하는 서비스를 수행하여야 할 것이다.

넷째, 연구자, 연구지원기관, 정부, 대학 및 대학도서관, 출판사 등 오픈 데이터와 관련 있는 모든 책임주체는 서로 협력하여 미해결 문제들을 차근차근 풀어가야 할 것이다. 일례로 IT의 발전이 오픈 데이터를 보다 확실하게 견인할 것이라고 전망하기 쉬운데, 비약적으로 발전한 IT가 오히려 오픈 데이터에 부담이 될 수 있다는 사실도 참조하여야 할 것이다. 사회 과학 연구에 많이 쓰는 Excel 파일은 20년 이내는 별 문제 없이 이용할 수 있을 것이나 그 이후에도 계속 잘 쓸 수 있을지를 담보하기 어렵다. 결과적으로 급속한 IT의 발달이 축적된 데이터의 계속 이용을 저해하지 않도록 데이터의 파일 형식, 구동 소프트웨어, 매체 전환 등에 관한 장기 계획을 잘 수립하여 추진할 필요가 있다. 또한 개인적으로 데이터를 관리하고 소장하는 대신 공공 리포지터리에 축적하는 것은 전용 서버에 축적되므로 안정적이고, 연구자 개인이 이직하는 경우 생길 수 있는 데이터 손

실 위험도 방지할 수 있으며, 다른 연구자의 이용도 확실하게 보장할 수 있기 때문이다(van den Bos, Jenny, & Wulff, 2014).

재정적 기술적 문제가 해결된다 해도 오픈 데이터는 라이선스, 윤리적 제약 등과 같은 여러 가지 넘어야할 장벽이 적지 않다. 그러나 오픈 데이터의 필요성과 효과에 대하여 이미 연구자, 연구지원기관, 대학 및 대학도서관, 출판사, 정부 등이 공감하고 있고 이미 오픈 데이터의 경험을 쌓아가고 있으므로, 이들 책임주체들이 적극 협력한다면 관련 문제도 차츰 해결되고 오픈 데이터도 곧 본 궤도에 오를 것으로 기대된다.

4. 결론

이 연구는 연구 중 생산된 연구데이터를 공개하고 공유하며 재이용하게 하는 오픈 데이터에 관한 것으로, 해외의 오픈 데이터 정책 추진 상황을 유형별로 파악하였다. 연구결과 해외의 학회, 비영리단체, 연구지원기관은 공공의 이익을 추구하는 차원에서 오픈 데이터를 추진하였고, 이른 곳은 2000년대 초중반부터 시작하였던 것으로 보인다. 해외의 주요 출판사에서는 연구자로부터 데이터를 넘겨받아 논문을 엄격하게 심사하는데 활용하고 추후 이를 공개하여 검증성을 확보한다는 차원에서 2010년 이후 오픈 데이터를 적극 추진하고 있는 것으로 나타났다. 구체적으로 'Nature' 학술지(2015)에는 매주 200편 이상의 논문이 투고되지만 이 가운데 최종적으로 출판되는 것은 약 8% 정도이므로 스스로 엄격하게 검증하고 심사할 수밖에

없는 상황이고, 이런 상황에서 저자에게 데이터 제출을 요구하는 것은 자연스러운 일로 보아야 할 것이다.

연구 결과 오픈 데이터를 보다 활발하게 전개하기 위해서는 해결하여야 할 과제가 적지 않지만 각 책임주체의 역할은 비교적 분명해 보인다. 연구자가 오픈 데이터에 적극 참여하게 하려면 오픈 데이터에 관한 이해를 높이는 동시에 데이터 인용의 맥락에서 데이터 공유가 자동적으로 수행되는 프레임워크를 설계하는 등 기술적인 개선 노력도 하여야 할 것이다. 연구지원기관은 연구자에게 DMP를 요청하고 대학 및 소속기관은 이를 관리하고 교육하며, 리포지터리와 협력하여 데이터 아카이빙 표준을 정하는 일을 도와야 할 것이다. 정부는 각 책임주체가 자신의 이해만을 주장하지 않고 오픈 데이터에 서로 협력할 수 있게 분위기를 조성하

고 이끌어 연구의 활성화, 학문의 성장, 산업의 발전을 기획하여야 할 것이다.

오픈 데이터는 연구자의 이목을 끄는 중요한 이슈인 것은 사실이지만 국내외를 막론하고 이에 대한 실천 경험이 일천하고 선행연구가 많지 않아 이 연구 결과가 국내 오픈 데이터를 활성화하는 구체적인 전략 수립에 이르지 못하는 한계가 있다. 과거 오픈 액세스가 그러했듯이 세계 주요국이 오픈 데이터에 참여하는 것은 시간의 문제일 뿐 이제 곧 나타날 현상으로 보아야 할 것이다. 따라서, 계속하여 오픈 데이터에 관하여 국내 학회, 연구지원기관, 연구자를 대상으로 경험적인 후속 연구가 활발하게 수행되기를 기대한다. 문헌정보학계에서도 오픈 데이터에 보다 많은 관심을 갖고 해당 전문가를 양성하는 등 보다 구체적인 노력을 기울여야 할 것으로 보인다.

참 고 문 헌

- 강희중 (2012). 21세기 핵심자원, 국가과학데이터 활용을 위한 정책 과제. STEPI Insight, (91), 1-21.
- 김문정, 김성희 (2015). 과학기술분야 연구자의 연구데이터 공유의 영향요인에 대한 연구. 한국문헌정보학회지, 49(2), 313-334. <http://dx.doi.org/10.4275/KSLIS.2015.49.2.313>
- 김선태, 한선화, 이태영, 김용 (2010). 과학데이터 보존 및 활용모델에 관한 연구. 한국비블리아학회지, 21(4), 81-93.
- 김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455.
- 김지현 (2013). 국외 정부연구비지원기관의 연구데이터 관리정책 분석. 한국문헌정보학회지, 47(3), 251-274. <http://dx.doi.org/10.4275/KSLIS.2013.47.3.251>
- 남궁용권, 남궁지영 (2008). 논문 작성의 이론과 실제. 파주: 양서원.
- 박성욱, 한선화 (2011). 국가 과학데이터센터 구축의 경제적 파급효과 분석. 정보관리연구, 42(3), 55-69. <http://dx.doi.org/10.1633/JIM.2011.42.3.055>

- 성진석 (2011). 연구관련 산출물 관리체계 구축. *KSLA Bulletin*, (3), 11-14.
- 신영란 (2011). 사회과학 데이터 아카이브 서비스. *KSLA Bulletin*, (3), 7-10.
- 신영란, 정연경 (2012). 국내 인문사회 연구데이터 아카이브의 개선방안에 관한 연구. *한국기록관리학회지*, 12(3), 93-115.
- 윤종민, 김규빈 (2013). 과학데이터에 관한 입법례와 관리정책 그리고 대응방안: 호주, 미국, 중국을 중심으로. *기술혁신학회지*, 16(1), 63-100.
- 이상환, 심원식 (2009). 과학데이터의 공유와 활용. *KISTI 지식리포트*, (5), 1-36.
- 한국과학기술정보연구원 (2012). 국가 과학데이터 공유·활용체제 구축. 서울: 교육과학기술부.
- Brown, S., Humphrey, C., Leahey, A., Webster, P., Austin, C., Born, A., & Mills, S. (2015). Guidelines for the deposit and preservation of research data in Canada. Retrieved from <http://www.rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Canada-2015.pdf>
- Callaghan, S. (2014). Open data in the earth and climate sciences. In: Moore, S. A.(ed.) *Issues in open research data*. (pp. 89-106). London: Ubiquity Press. <http://dx.doi.org/10.5334/ban.f>
- Caulfield, T., Harmon, S. H. E., & Joly, Y. (2012). Open science versus commercialization: A modern research conflict. *Genome Medicine*, 4(17). DOI: 10.1186/gm316
- Dimitrova, V. (2014). Open research data in economics. In: Moore, S.A.(ed.) *Issues in open research data*. (pp. 141-150). London: Ubiquity Press. <http://dx.doi.org/10.5334/ban.i>
- Kansa, E. C., & Kansa, W. S. (2013). We all know that a 14 is a sheep: Data publication and professionalism in archaeological communication. *Journal of Eastern Mediterranean Archaeology and Heritage Studies*, 1(1), 88-97. Retrieved from http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/journal_of_eastern_mediterranean_archaeology_and_heritage_studies/v001/1.1.kansa01.html
- Karunakara, U. (2014). Data sharing in a humanitarian organization: The experience of Medecins Sans Frontieres. In: Moore, S.A. (ed.) *Issues in open research data*. (pp. 59-76). London: Ubiquity Press. <http://dx.doi.org/10.5334/ban.f>
- Kim, Jeonghyun (2013). Data sharing and its implications for academic libraries. *New Library World*, 114(11/12), 494-506. DOI: 10.1108/NLW-06-2013-0051
- KOSSDA (2015). Retrieved from http://www.kosdda.or.kr/w01_03a.asp
- Kratz, J., & Strasser, C. (2014). Data publication consensus and controversies. <http://dx.doi.org/10.12688/f1000research.3979.2>
- Mounce, R. (2014). Open data and palaeontology. In: Moore, S.A. (ed.) *Issues in open research*

- data. (pp. 151-164). London: Ubiquity Press. Retrieved from <http://dx.doi.org/10.5334/ban.j>
- Murphy, F. (2014). Data and scholarly publishing: The transforming landscape. Learned Publishing, 27(special issue), S3-S7. DOI: 10.1087/20140502
- Nature Publishing Group (2013). Availability of data and materials. Retrieved from <http://www.nature.com/authors/policies/availability.html>
- Nature Publishing Group (2015). Getting published in Nature: The editorial process. Retrieved from http://www.nature.com/nature/authors/get_published/index.html#a1
- Norman, H. (2014). Mandating data archiving: Experiences from the frontline. Learned Publishing, 27(special issue): S35-S38. DOI: 10.1087/20140507
- PLOS (2013). Data access for the open access literature: PLOS's data policy. Retrieved from <https://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
- Pollard, T., & Celi, L. A. (2014). Open data in health care. In: Moore, S.A. (ed.) Issues in open research data. (pp. 129-140). London: Ubiquity Press. <http://dx.doi.org/10.5334/ban.h>
- Research Councils UK (2013). RCUK policy on open access and supporting guidance. Retrieved from <http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf>
- Royal Society (2012). Science as an open enterprise. The Royal Society Science Policy Centre Report 02/12. Retrieved from http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf
- van den Bos, W, Mirjam, J., & Wulff, D. (2014). Open minded psychology. In: Moore, S (ed.) Issues in open research data. (pp. 107-127). London: Ubiquity Press. <http://dx.doi.org/10.5334/ban.g>
- Vines, T. H. et al. (2013). Mandated data archiving greatly improves access to research data, The FASEB Journal, 27, 1304-1308. DOI: 10.1096/fj.12-218164
- The White House OSTP (2013). Expanding public access to the results of federally funded research. Retrieved from <https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Kang, Hee-Jong (2012). Key resources in 21st century, policy tasks on use of national scientific

- data. STEPI Insight, (91), 1-21.
- Kim, Jihyun (2012). A study on university researchers' data management practices. *Journal of 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. *Journal of the Korean Society for Library and Information Science*, 47(3), 251-274. <http://dx.doi.org/10.4275/KSLIS.2013.47.3.251>
- Kim, Moonjeong, & Kim, Seonghee (2015). A study on the factors affecting sharing of research data of science and technology researchers. *Journal of the Korean Society for Library and Information Science*, 49(2), 313-334. <http://dx.doi.org/10.4275/KSLIS.2015.49.2.313>
- Kim, Sun-Tae, hahn, Sun-Hwa, Lee, Tae-young, & Kim, Yong (2010). A study on a model for using and preserving scientific data. *Journal of the Korean Biblia Society for Library and Information Science*, 21(4), 81-93.
- Korean Institute of Science and Technology Information (2012). A development on the scientific data sharing and utilizing system. Seoul: Ministry of Education.
- Lee, Sang-Hwan, & Shim, Wonsik (2009). Sharing and use of scientific data. *KISTI Knowledge Report*, (5), 1-36.
- Namkung, Yong-Kwon, & Namkung, Ji-young (2008). Theory and practice on writing papers. Paju: Yangseowon.
- Park, Sung-Uk, & Hahn, Sun-Hwa (2011). An economic ripple effect analysis of national scientific data center construction. *Journal of Information Management*, 42(3), 55-69. <http://dx.doi.org/10.1633/JIM.2011.42.3.055>
- Shin, Young-Ran (2011). Data archive service in social science. *KSLA Bulletin*, (3), 7-10.
- Shin, Young-Ran, & Chung, Yeon-Kyoung (2012). A study on the improvement plans of the humanities and social sciences research data archives in Korea. *Journal of Korean Society of Archives and Records Management*, 12(3), 93-115.
- Sung, Jin-Seok (2011). Development of management system for archiving research data. *KSLA Bulletin*, (3), 11-14.
- Yoon, Chong-min, & Kim, Kyubin (2013). Legislation cases, management policies and countermeasures on scientific data: Focusing Australia, the United States and China. *Journal of Korea Technology Innovation Society*, 16(1), 63-100.