

데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구*

A Study on the Perceptions of University Researchers on Data Management and Sharing

김 지 현 (Jihyun Kim)**

목 차

- | | |
|--------------------------|-----------------|
| 1. 서 론 | 3.3 데이터 저장 및 보존 |
| 2. 대학 연구자들의 연구데이터 관리와 공유 | 3.4 데이터 공유 |
| 3. 연구방법 및 인터뷰데이터 분석 | 3.5 데이터 재이용 |
| 3.1 데이터 생성 및 수집 | 4. 요약 및 제언 |
| 3.2 데이터 기록화 | 5. 결 론 |

초 록

본 연구는 국내 대학에 소속된 연구자들을 대상으로 데이터 관리 현황과 데이터 공유 및 재이용에 대한 경험과 인식을 조사하는 것을 목적으로 하였다. 이를 위해 본 연구에 선행하여 수행된 설문조사의 응답자 중 후속 인터뷰에 동의한 13명을 대상으로 반구조화된 인터뷰를 수행하였다. 참여자들은 다양한 유형과 포맷의 데이터를 생성 또는 수집하고 있었으며 데이터 기록화를 수행하는 연구자들은 소수에 불과하였으나 이들은 그 중요성을 인식하고 있었다. 데이터가 유용한 기간을 데이터가 논문 출판에 활용될 수 있는 기간으로 인식하는 연구자들이 대다수이었으나, 데이터가 유용하다고 인식하는 기간 이상으로 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. 정부연구비 지원을 받는 과제의 데이터는 오픈 액세스로 공개해야 한다는 논리에 전적으로 찬성하는 연구자들도 있는 반면 부분적으로 동의하거나 반대하는 의견도 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.

ABSTRACT

This study aimed to investigate data management practices of university researchers in Korea, as well as their experiences and perceptions of data sharing and reuse. For this purpose, it performed semi-structured interviews of 13 researchers who agreed to participate in interviews followed by a survey conducted prior to this study. The interview participants created or collected research data with various types and formats, and only a few conducted data documentation while they recognized its significance. The majority of participants perceived the period that data would be useful as the period that data can be employed for publications. However, most participants wanted to preserve data beyond the period that data would be considered useful and it indicates they have no small need for data storage and preservation. Participants usually shared data with those whom they have known, such as a personal research group or a research team. While some completely agree with the principle that publicly-funded data should be open to the public, others partially agreed or disagreed with it. Most participants were concerned about being scooped, plagiarism, and maintaining the first right to publish and incentives to mitigate the concerns would be necessary.

키워드: 데이터관리, 데이터공유, 데이터재이용, 대학 연구자

Data Management, Data Sharing, Data Reuse, University Researchers

* 이 논문은 2012년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 연구되었음 (NRF-2012S1A5A8022036).

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

논문접수일자: 2015년 7월 16일 최초심사일자: 2015년 7월 21일 게재확정일자: 2015년 8월 4일
한국문헌정보학회지, 49(3): 413-436, 2015. (<http://dx.doi.org/10.4275/KSLIS.2015.49.3.413>)

1. 서론

학술 커뮤니케이션에서 연구데이터가 가지는 잠재적 가치와 데이터의 효과적인 관리 및 공유에 대한 관심은 그 어느 때보다 높다. 북미와 유럽 선진국들에서는 학술 연구를 통해 생성된 데이터를 그 생애주기에 따라 적극적·지속적으로 관리하는 활동을 ‘데이터 큐레이션(data curation)’으로 정의하고 이에 포함되는 세부 활동과 실무에 대한 연구를 활발히 진행하고 있다. 이를 통해 데이터의 검색, 데이터 품질 유지, 부가가치 창출, 데이터의 재이용 촉진을 강화할 수 있으므로 해외에서는 데이터센터 및 데이터 레포지터리, 대학 내 기관 레포지터리 등 데이터 관리 주체들을 중심으로 데이터 큐레이션을 적극적으로 실행하고 있다.

이와 더불어 연구데이터를 오픈액세스로 공유하려는 활동 역시 활발하게 전개되고 있다. 예를 들어 미국 연구도서관협회(ARL: Association of Research Libraries)에서는 최근 ‘학술연구 공개(Open Scholarship)’라는 핵심 영역 아래 오픈액세스, 데이터 공개(Open Data), 그리고 교육자료의 공개(Open Educational Resources)를 제시하고 있다. 이 중에서 데이터 공개는 데이터의 일부 혹은 그 내용을 누구나 이용, 재이용 또는 재배포할 수 있게 하는 것을 의미한다(ARL 2014). 또한 영국을 중심으로 하는 비영리기관인 Open Knowledge Foundation의 경우 연구데이터 공개를 위한 실무 그룹을 결성하여 영국 캠브리지 대학 교수들과의 협력을 통해 연구데이터 공개 원칙(Panton Principles for Open Data in Science)을 제안하였다. 이 원칙에서는 데이터 제공자가 데이터를 공개할

때 원하는 이용 조건을 상세히 명시할 것과 데이터 공개에 적합한 라이선스를 사용할 것 등 실제적인 데이터 공개 방안을 제시하고 있다(Molloy 2011). 이러한 비영리기관의 활동뿐만 아니라 정부연구비를 제공하는 미국과 영국 등지의 기관에서는 지원받은 과제에서 생산되는 데이터의 공개를 의무화하는 정책을 수행하고 있다(Goben and Solo 2013). 이와 같이 해외에서는 연구데이터와 관련된 다양한 주체들이 데이터의 공유 및 공개 활동에 참여하고 있음을 알 수 있다.

연구데이터 관리와 공유에 있어서 데이터를 생산하는 연구자들의 긍정적인 인식과 참여가 매우 중요한 요소라고 할 수 있다. 실제로 해외에서는 많은 관련 연구들이 데이터 생산자의 관점에서 수행되어 왔는데 Borgman(2015)은 데이터를 관리하는 방법은 학문 분야마다 매우 다양하고, 연구자들을 데이터 공유에 참여시키는 것은 어려운 일이며 이를 위한 효과적인 유인책도 부족한 상황임을 지적하였다. 국내에서는 정부출연연구기관 연구자들을 대상으로 데이터의 관리 및 공유에 대한 인식을 조사한 연구가 진행되었고 모두 설문조사를 통해 연구데이터 공유에 영향을 미치는 요인에 대한 분석을 실시하였다(김은정 2012; 김문정, 김성희 2015). 국내에서도 데이터 관리와 공유에 대한 관심이 높아지고 있는 상황이므로 이를 활성화하기 위한 기초 연구로서 다양한 연구 분야 및 기관에서의 연구자 인식과 경험에 대한 연구가 필요할 것이다.

따라서 본 연구에서는 대학에 소속된 연구자 13명을 대상으로 반구조화된 인터뷰(semi-structured interview)를 통해 그들이 생산하는 연구데이

터의 관리 현황과 공유 및 재이용에 대한 경험과 인식을 조사하였다. 국내의 연구자들을 대상으로 수행된 연구가 소수에 불과하고 국내 연구자들 중 많은 수가 대학에 소속되어 있음을 감안할 때 대학 연구자들의 데이터 관리와 공유에 대한 인식을 조사하는 것이 필요하다. 본 연구에서는 인터뷰를 통한 질적 연구를 통해 연구자들의 구체적인 의견과 경험을 살펴보고자 하였다. 연구데이터의 관리와 공유의 제도적 체계가 미흡한 국내 현실에서 대학 연구자들이 인식 혹은 경험한 데이터 관리 및 공유의 문제점과 한계를 논의하고 이를 촉진할 수 있는 기초를 마련하는데 본 연구의 의의가 있다.

2. 대학 연구자들의 연구데이터 관리와 공유

해외에서는 연구데이터의 관리 및 공유와 관련된 문제점 혹은 연구자들이 수행하고 있는 모범사례를 조사하기 위해 대학 연구자들의 인식과 경험을 조사하는 다양한 연구들이 진행되어 왔다. 선행 연구는 크게 대학 연구자들의 데이터 관리와 공유 현황과 이에 대한 인식 전반을 조사하여 향후 대학도서관 연구데이터 관리 서비스의 기초 자료로 활용하고자 하는 연구들과 데이터 관련 활동 중에서도 데이터 공유에 대한 경험과 인식에 초점을 맞추어 진행한 연구들로 구분할 수 있다.

데이터 관리와 공유 전반에 대한 조사 연구로서 Scaramozzino et al.(2012)는 캘리포니아의 강의중심 대학 한 곳의 교수 및 강사를 대상으로 설문조사를 수행하였다. 응답자들은 메타

데이터 생성, 데이터 보존 및 공유, 데이터관리 계획 작성, 데이터 관리 모범사례에 대한 추가적인 안내와 교육을 원하고 있는 것으로 나타났다. 그러나 대학도서관이 데이터 관리를 전문적으로 수행하는 기관이라는 인식은 낮은 것으로 나타나 교수와 사서 간 적절한 협력관계의 모색이 필요하다고 보았다.

Westra(2010)는 미국 오레곤 대학에서의 데이터 서비스 수립에 앞서 서비스에 대한 요구 조사를 위해 대학 내 연구자들을 대상으로 온라인 설문조사 및 인터뷰를 수행하였다. 인터뷰 참여자들은 데이터 저장 및 백업을 위한 인프라 제공과 데이터 관리를 위한 도구 제공을 데이터 관리의 주요 이슈로 제시하였다. 참여자들의 데이터는 주로 PC, 노트북, 외장하드, 광디스크 등에 저장되고 있었으나 참여자들이 복수의 저장 공간을 활용하여 데이터를 저장하는 경우는 드문 것으로 나타났다. 다수의 연구자들은 데이터의 생애주기에 맞게 데이터 생성 및 획득 단계에서부터 관리를 도와주는 도구가 필요하다는 의견을 제시하였다.

최근에는 영국의 대학들을 중심으로 기관 내에서 수행되는 데이터 관리 현황을 조사하고 데이터 서비스 구축에 필요한 기초 자료를 수집하기 위한 방법으로서 데이터 자산 프레임워크(Data Asset Framework, 이하 DAF)라는 설문도구가 활용되고 있다. DAF는 기관의 경영진이나 연구비지원기관, 연구자, 데이터 레포지터리 실무자, IT 서비스 등 네 유형의 이해관계자를 대상으로 개발된 설문 내용을 포함하고 있다. 이 중에서도 영국의 에든버러(Edinburgh), 노팅햄(Nottingham), 엑스터(Exeter), 노스햄프턴(Northampton) 대학에서는 연구자 대상

설문도구를 활용하여 각 대학 소속 연구자들을 대상으로 데이터 관리 현황 조사를 실시한 사례가 있다. 이는 데이터 관리와 공유에 대한 대학 연구자들의 요구를 공통된 설문도구를 활용하여 분석을 수행하였다는 점에서 주목할 만한 시도라고 할 수 있다.

DAF를 활용하여 대학 연구자들의 인식을 조사한 연구들을 살펴보면 우선 영국 노팅햄 대학(University of Nottingham)의 연구자 366명의 설문 결과를 통해 다양한 유형의 데이터가 생산, 수집되며 연구노트와 같은 물리적 형태의 자료도 중요한 데이터로 인식되고 있음을 알 수 있었다. 데이터는 대학에서 제공하는 서버뿐만 아니라 외장하드, USB, Dropbox와 같은 외부 서비스에도 저장되고 있었다. 대학 외부의 연구자들과 데이터를 공유하는 경우는 드물었으며 개인정보 등 민감한 정보를 담은 데이터의 공유는 꺼리는 것으로 나타났다(Parsons et al. 2013).

영국 엑서터 대학의 연구자들 역시 다양한 저장 장치를 데이터 저장에 활용하고 있었으며 다수의 연구자들이 20GB의 네트워크 저장 공간을 데이터 백업 용도로 필요하다고 응답하였다. 연구자들은 논문과 마찬가지로 데이터에 대한 적절한 인용이 필요하다고 인식하고 있었다(Open Exeter Project Team 2012).

노스햄튼 대학에서는 소속기관 연구자들을 대상으로 설문과 인터뷰를 수행하였으며 이들이 생산하는 데이터의 포맷은 문서, 스프레드시트, 이미지, 오디오, 비디오, 데이터베이스 등 다양하였다. 연구자들은 수집한 데이터를 대학 내 서버, 노트북 컴퓨터 등에 주로 저장하고 이를 이메일, USB, 외장하드, 대학에서 제

공하는 공유 드라이브 등에 백업하는 것으로 나타났다. 연구가 종료된 이후에는 일반적으로 CD나 DVD 또는 외장하드에 데이터를 저장하고 있었다. 데이터 공유는 같은 연구팀 내에서만 이루어지고 있었으며 오픈 액세스로 데이터를 공유하는 것에 대한 우선순위는 낮은 것으로 나타났다(Alexogiannopoulos et al. 2010).

에든버러 대학의 경우 DAF를 활용하여 대학 내 연구센터 및 특정분야 연구자들을 대상으로 사례조사를 실시하였다. 연구자들은 공통적으로 데이터의 원활한 저장과 활용을 위해 대학에서 제공하는 서버 저장 공간의 확대를 요청하고 있었다. 또한 연구자들이 생산하는 데이터의 대부분은 유실되었을 경우 재생산이 불가능한 유형이었으므로 적어도 10년의 보유 기간을 원하고 있었다. 데이터의 생성 및 저장에 있어 표준화된 절차와 지침에 대한 연구자들의 요구가 높은 것으로 나타났다(Ekmekcioglu and Rice 2009).

데이터와 관련된 대학 연구자들의 활동 중에서도 데이터 공유에 초점을 맞추어 수행된 연구들도 다수 존재한다. 소규모 과학 분야(Little science) 연구자들의 데이터 관리에 대한 연구로서 Borgman et al.(2007)는 거대과학 분야(Big science)에서 데이터 레포지토리를 구축하여 데이터를 관리하는 경우가 많은 반면 소규모 과학 분야에서는 연구자가 속한 그룹 단위로 데이터가 관리되고 있음을 언급하였다. 소규모 과학에 해당하는 서식생태학(habitat ecology) 분야 연구자들을 대상으로 인터뷰와 관찰을 기반으로 하는 질적 연구를 수행한 결과 데이터 공유가 중요하다는 인식은 존재하지만 실제로는 거의 공유가 이루어지지 않는다는 사실을 확

인하였다. 이와 관련된 요인으로는 데이터에 대한 요청 부족, 적합한 표준 부재, 논문 출판에 대한 우려, 데이터 소유권의 문제, 데이터의 품질, 연구윤리 측면에서의 우려 등이 있었다.

Wallis et al.(2013)의 연구는 10년 동안 이틀 서식생태학자들이 속한 연구센터인 Center for Embedded Network Sensing(CENS)의 데이터 관리 현황과 연구자들의 인식을 조사한 내용을 바탕으로 하고 있다. 데이터의 수집과 보존에 있어 거대과학 분야에서처럼 정립된 계획에 따라 고가의 장비를 활용하여 수집하여 보존하는 것이 아니라 소규모 과제 단위로 몇 명의 연구자들에 의해 수집이 이루어지며 보존이나 재이용은 거의 이루어지지 않는다는 사실을 제시하였다. CENS에 속한 연구자들은 그들의 데이터를 오픈 액세스로 누구나 활용할 수 있게 공유하기보다는 개인적으로 잘 알고 신뢰하는 연구자들을 대상으로 공유하는 경향을 보였다. 이는 학술연구에서의 나눔의 문화(gift culture in scholarship)가 신뢰에 기반을 둔 관계 안에서 개인적인 교환의 형식으로 이루어지고 있는 현상을 반영하는 것으로 보였다. 연구자들은 데이터를 활용하여 최초로 논문을 출판할 수 있는 권리를 보장받는 것과 데이터를 공유했을 경우 이를 재이용하는 연구자들이 데이터 제공자에 대한 사사표기 및 인용을 제공하는 것을 중시하는 것으로 나타났다.

Cragin et al.(2010) 역시 소규모 과학 분야의 연구자 20명을 대상으로 데이터 공유 및 보존에 대한 인식을 조사하였다. 연구자들은 대체적으로 데이터 공유에 대해 긍정적인 인식을 가지고 있었으나 60%에 해당하는 12명의 연구자들은 데이터로부터 모든 논문을 출판한 이후

에 공유를 할 것이라고 응답하였다. 나머지 8명의 연구자들도 논문을 출판하기 이전에 데이터를 공유한다면 잘 알고 있고 신뢰하는 연구자들과 공유하는 것으로 하여 데이터 공유의 범위를 철저히 제한하고 있었다. 또한 데이터 공유를 통해 공저자가 되는 것에 대하여 대부분의 연구자들이 그 필요성을 강조하였다. 데이터의 보존 기함에 있어 절반 이상인 65%의 연구자들이 10년 이상을 원하였으며 20%의 연구자들이 5-10년이라고 응답하였다. 장기적인 보존 가치를 가지는 데이터는 주로 비교 연구의 가치가 있거나 장기적인 종단 연구에 활용될 가능성이 있는 관찰 데이터인 것으로 나타났다.

Akers와 Doty(2013)는 에모리 대학 교수 330명의 설문 응답을 바탕으로 자연과학, 사회과학, 의학, 예술·인문학 분야 교수들 간 데이터 관리 및 공유의 차이점을 분석하였다. 자연과학분야 교수들의 경우 가장 많은 양의 디지털 데이터를 보유하고 있었으며 대학에서 제공하는 서버에 데이터를 저장하면서 실험에서 사용하는 특수한 장비 내에 있는 하드 드라이브를 데이터 저장 공간으로 활용하고 있었다. 또한 데이터와 메타데이터에 대한 기록을 연구노트 및 수기로 작성한 문서로 남기는 경우가 많아 물리적 데이터와 디지털 데이터의 통합에 어려움을 겪는 것으로 나타났다. 이들은 연구비지원기관에서 요구하는 데이터관리계획에 대한 인식도가 높았으며 외부 연구자들과도 데이터뱅크나 레포지토리를 통해 비교적 활발히 데이터를 공유하고 있었다. 이들이 데이터 공유를 꺼리는 가장 큰 요인은 데이터에 대한 적절한 인용을 받지 못할 것이라는 우려인 것으로 나타났다. 이에 비해 사회과학분야 교수들은 데

이터 공유에 있어 인간 피험자 관련 개인정보 보호와 비밀유지를 가장 중시하는 것으로 나타났다. 이러한 결과는 의학 분야 교수들의 인식과도 부합하였다. 예술·인문학 분야 교수들의 경우 연구의 기반이 되는 원 자료가 다른 분야에서 통용되는 '데이터'의 개념보다 포괄적이므로 데이터의 양을 추정하는데 어려움을 느끼는 것으로 나타났다. 이들은 물리적 자료의 디지털화 및 일반 대중들과의 데이터 공유에 대해 다른 분야 교수들보다 높은 관심을 보였다.

Williams(2013)는 데이터를 오픈 액세스로 공유하고 있는 작물학(crop science) 분야 교수 7명을 대상으로 인터뷰를 수행하였다. 이들은 연구 분야에서의 데이터 공유에 대한 기대, 다른 연구자들을 도울 수 있다는 점, 그리고 연구비지원기관의 요구사항이기 때문에 데이터를 공개하고 있다고 응답하였다. 이들은 분야별 데이터 레포지토리를 통해 데이터를 공유하고 있었는데 논문 작성 및 연구 프로포절 제출 등 우선순위가 높은 업무와 데이터 공유 작업 간 균형을 맞추는 것이 쉽지 않다는 점을 언급하였다. 대학도서관이 데이터 관리와 공유를 지원함으로써 그 역할이 확장될 수 있다는 관점에 대해 인터뷰에 참여한 교수들은 대부분 그렇게 생각해본 적은 없으나 그 필요성과 실현 가능성을 높이 평가하는 것으로 나타났다.

Whyte과 Pryor(2011)은 천문학, 생물정보학(bioinformatics), 화학, 전염병학, 언어학, 신경영상 분야 연구자들을 대상으로 수행한 인터뷰를 통해 데이터의 공개(openness)에 대한 인식을 조사하였다. 인터뷰 참여자들은 데이터 수집에 드는 비용이 절약되고 협동 연구가 강화될 수 있다는 점에서 연구의 효율성이 증진될

수 있으며 연구 질문과 분석 방법의 발견 등 새로운 연구 역량의 강화, 연구의 가시성이 높아지고 연구결과의 상업화 기회도 확대될 수 있으므로 궁극적으로 연구의 영향력이 강화될 수 있다는 것을 장점으로 들었다. 그러나 데이터를 어느 수준에서 공개할 수 있는지에 관해서는 연구자들마다 다르게 인식하고 있는 것을 나타냈다. 이들은 자신이 속한 연구그룹 내에서의 공유, 협동 과제를 수행하는 파트너들 간의 공유, 같은 분야 동료들 간 조건부 공유, 자신이 속한 연구 커뮤니티 내에서의 제한적인 공유, 누구나 활용할 수 있게 공개하는 방법 등 다양한 방식의 데이터 공유 수준을 언급하고 있었다. 연구자들의 데이터 공유를 촉진하기 위해서는 데이터 제공자의 기여도를 인정하고 협력 연구의 기회를 높이며 재이용할 수 있는 데이터를 확대하는 것이 필요하며 데이터 제공에 드는 비용과 이로 인한 편익에 대한 면밀한 분석이 요청된다는 것을 제안하였다.

Kim and Stanton(2011)은 미국 동부에 위치한 대학 3곳에서 과학, 기술, 공학 및 수학(STEM: Science, Technology, Engineering and Mathematics) 분야의 연구자들을 대상으로 데이터 공유 현황과 영향 요인을 조사하는 인터뷰를 수행하였다. 이 연구에서는 25명의 연구자들과의 인터뷰 데이터를 분석하여 데이터 공유에 영향을 미치는 요인을 크게 9가지로 구분하였다. 이를 살펴보면 첫째, 연구비지원기관이나 학술지에서 데이터 공유를 의무화하는 등의 강제적 요인, 둘째, 데이터 공유가 전문가로서의 사명이나 동료 연구자들이 기대하는 행동으로 인식되는 규범적 요인, 셋째, 다른 연구자들이 공유된 데이터를 활용하여 연구 역량을

높이는 사례를 관찰하고 이를 따라하게 되는 모방적 요인, 넷째, 연구의 품질을 인정받고 연구자로서의 명성을 높이는 등 데이터 공유의 장점에 대한 인식, 다섯째, 데이터 공유를 실행하기 위해 투자해야 하는 노력이나 시간 등 비용적 측면에 대한 인식, 여섯째, 논문을 출판할 기회가 줄어들 수도 있다는 우려 등 공유에 따른 위험에 대한 인식, 일곱째, 연구자 스스로 가지고 있는 IT 기술 활용 능력, 여덟째, 데이터 레포지터리나 메타데이터 표준 등 외부에서 제공하는 IT 능력, 아홉째, 학문의 발전에 기여하고 다른 이들의 연구를 도울 수 있다는 이타적인 동기를 포함한다.

선행연구를 통해 대학 연구자들은 다양한 유형과 포맷의 데이터를 생성하며 이를 저장 또는 보존하기 위해 여러 종류의 저장 공간과 Dropbox와 같은 외부 서비스를 활용하고 있었다. 대부분의 조사 연구에서 데이터의 저장 및 보존 서비스에 대한 연구자들의 요구가 높은 것으로 나타났다. 또한 연구자들은 대학도서관을 데이터 관리의 주체로 인식하는 경향은 낮았으나 그 가능성에 대해서는 긍정적으로 평가하였다. 데이터 공유에 있어 대다수의 연구자들이 공유로 인한 비용적인 요인 또는 위험 요인에 대한 우려를 가지고 있었다. 데이터 공유에 대한 인식과 필요로 하는 서비스에 대한 요구 역시 연구 분야 별로 차이가 있음을 확인할 수 있었다.

3. 연구방법 및 인터뷰데이터 분석

본 연구에서는 대학 연구자들을 대상으로 설

문조사를 수행한 후 설문 응답자들 중 후속 인터뷰에 대한 이메일을 수신하는데 동의한 연구자들 중에서 인터뷰 참여자를 모집하였다. 이를 통해 이루어진 후속 인터뷰 내용에 초점을 맞추어 본 연구의 분석을 수행하였다.

인터뷰 참여자들이 설문조사의 응답자이므로 이들이 포함되었던 설문조사 모집단과 표본의 추출 방법을 간략히 설명하고자 한다. 설문조사의 모집단은 2010년과 2011년에 한국연구재단에서 최종 선정된 과제의 연구책임자 12,984명으로 하였다. 이들은 국가연구비를 지원받은 연구자들이므로 그렇지 않은 연구자들에 비해 연구데이터 관리와 공유에 있어 유리한 입장에 있다고 판단하였다. 이 중에서 무작위 추출을 통해 10%에 해당하는 인원을 선정한 후 중복되는 연구자와 온라인 설문조사 도구로 활용하였던 서베이몽키 이메일을 수신 거부하는 연구자들을 제외하였다. 그 결과 인문·사회 분야 342명, 자연과학·공학 분야 675명을 포함하여 총 1,017명의 연구자들을 표본으로 선정하였으며 이들에게 온라인 설문지를 송부하였다.

설문조사는 2012년 8월 2일부터 9월 3일까지 실시하였으며 192명(18.9%)의 설문 응답을 회수하였다. 이 중에서 후속 인터뷰에 대한 이메일 수신에 동의한 응답자는 65명이었고 이들을 대상으로 인터뷰 참여에 대한 의사를 문의하는 이메일을 송부하였다. 그 결과 인터뷰 참여에 최종적으로 동의한 응답자는 13명이었으며 이들을 대상으로 2012년 11월 5일부터 12월 12일 사이에 전화 인터뷰를 수행하였다.

〈표 1〉에서와 같이 인터뷰 참여자들은 1명의 박사후 연구원과 1명의 연구교수를 제외하고는 모두 정년 트랙 교수들이었다. 연구 분야

〈표 1〉 인터뷰 참여자의 특성 및 생성/수집하는 데이터 유형

ID	직위	연구 분야	성별	연령대	생성/수집하는 데이터 유형
P3	정교수	관광경영학	남	50대	관찰데이터
P16	조교수	불교학	남	40대	관찰데이터, 기록
P31	조교수	행정학	남	30대	관찰데이터
P36	조교수	경영학	남	30대	관찰데이터
P37	조교수	경영학	남	30대	관찰데이터
P41	조교수	통계학	남	30대	계산데이터
P45	부교수	법학	여	40대	관찰데이터, 기록
P47	부교수	컴퓨터공학	남	30대	계산데이터
P49	부교수	토목공학	남	30대	계산데이터
P50	박사후연구원	지질학	여	20대	실험데이터
P51	연구교수	정치외교학	남	40대	관찰데이터
P59	정교수	종교학	남	40대	기록
P62	조교수	한의학	남	30대	실험데이터

별로는 컴퓨터공학, 토목공학, 통계학, 지질학, 한의학 등 이공 및 의학 분야 연구자 5명, 경영학, 행정학, 법학, 정치외교학 등 사회과학 분야 연구자 6명, 그리고 불교학, 종교학 등 인문학 분야 연구자 2명으로 구성되었다. 2명을 제외하고는 남성 연구자들이었으며 연령대별로 20대 1명, 30대 7명, 40대 4명, 50대 1명으로 대부분 30대-40대 연구자들이 인터뷰에 참여한 것으로 나타났다. 본 연구에서 제시한 연령과 직위는 인터뷰를 수행한 2012년을 기준으로 작성되었다.

인터뷰는 준비한 인터뷰질문지의 질문을 토대로 필요할 경우 새로운 질문을 추가하는 반구조화된 방식으로 이루어졌다. 각 참여자 별로 인터뷰에 약 35분에서 50분 정도의 시간이 소요되었으며 참여자의 동의하에 인터뷰 내용을 녹취하여 분석을 실시하였다. 인터뷰 데이터의 분석을 위해 선행연구를 토대로 코딩체계를 개발하였으며 (1) 데이터 유형; (2) 데이터 수집방법; (3) 데이터 기술; (4) 데이터 저장 및 보

존; (5) 데이터 공유 및 재이용을 상위 코드로 하여 구성하였다.

분석을 통해 각 참여자들이 생성 혹은 수집하는 데이터 유형과 실험, 관찰, 시뮬레이션, 설문, 인터뷰 등 다양한 데이터 수집 및 생성 방법에 대하여 살펴보았다. 또한 데이터에 대한 메타데이터를 제공하거나 설명을 제공하는 등 데이터 기록화에 대한 인식 및 현황을 논의하였으며, 데이터의 저장 및 보존 기한에 대한 인식, 데이터의 공유와 재이용 현황과 이에 영향을 미치는 요인에 대해서도 조사하였다.

3.1 데이터 생성 및 수집

인터뷰 참여자들은 분야마다 다양한 종류의 연구데이터를 생성 또는 수집하고 있었다. Borgman(2012)에 따르면 연구데이터의 유형은 크게 관찰(observational) 데이터, 계산(computational) 데이터, 실험(experimental) 데이터, 기록(records)으로 구분할 수 있다. 관

찰 데이터는 설문 또는 측정 등 다양한 방법으로 생성될 수 있으며 구체적인 혹은 다수의 시공간과 관련된 정보를 제공하는 데이터를 의미한다. 계산 데이터는 컴퓨터 모형 혹은 시뮬레이션의 결과로 생성되는 데이터를 말하며, 실험 데이터는 실험실 환경 또는 현장실험에서 생성된 데이터를 의미한다. 기록은 정부, 민간기업 등 다양한 주체 및 기관에서 생성되는 원 자료로서 과학 연구뿐만 아니라 인문 사회과학 연구에 필요한 데이터를 제공한다. 인터뷰참여자들이 주로 생성 또는 수집한다고 응답한 데이터의 유형은 <표 1>에 제시되어 있다.

이공 및 의학 분야 참여자들의 경우 주로 계산 데이터나 실험 데이터를 생성하고 있었는데 계산 데이터의 예로 토목공학 분야 연구자(P49)의 경우 컴퓨터로 지진이 발생했을 때 구조물이 어떻게 진동을 하는지를 시뮬레이션을 통해 계산하고 그 해석결과가 수치화된 형태로 생성됨으로써 연구데이터를 얻게 된다고 설명하였다. 통계학 연구자(P41)의 경우 주로 외국에서 오픈 액세스로 공개된 데이터를 분석에 활용하는데, 이를 분석하기 위해서는 시뮬레이션에 필요한 함수와 이를 실행하기 위한 코드를 개발해야 하기 때문에 이에 많은 노력을 기울인다고 설명하였다. 따라서 연구자 본인이 개발한 시뮬레이션 함수와 코드가 연구데이터의 핵심이라고 설명하였다. 비슷한 경우로 컴퓨터공학 분야 연구자(P47) 역시 효율적인 무선통신을 위한 알고리즘을 개발하는 연구를 수행하면서 이를 실행하기 위한 소스코드가 중요한 데이터임을 언급하였다.

논문을 쓰다 보면 당연히 거기에서 그런 알고리

즘을 개발하기 위한 소스코드, 소위 말해 컴퓨터 프로그램[.....] 대부분 다 컴퓨터 프로그램을 돌려서 그 결과가 좋게 나와서 우리의 알고리즘이 좋다는 것을 입증하는, 그러한 것이거든요. 그러니까 이렇게 하면 좋다, 왜 좋냐, 이렇게 컴퓨터 모의실험을 통해서 보여줬다, 라고 해서 [.....] 개발한 소스코드는 잘 보관을 해야 합니다. [.....] 그 데이터, 다시 말해 소스코드, 컴퓨터 프로그램이 굉장히 중요하기 때문에. (P47, 컴퓨터공학, 부교수, 30대)

실험데이터를 주로 생산한다고 응답한 지구과학 분야 연구자(P50)는 암석표본을 채취한 뒤 이에 대한 지화학적 분석을 실시하여 구성 광물과 동위원소 분석 등의 결과를 담은 데이터를 주로 다루고 있다고 설명하였다. 한의학 분야 연구자(P62)의 경우에는 실험데이터를 세 가지 유형으로 설명하였는데 실험 장비를 사용하여 생성되는 수치화된 데이터, 실험대상 생물의 조직을 염색하여 관찰할 수 있게 하는 영상 디지털 데이터, 그리고 사람이 직접 관찰하여 수치를 측정 후 기록한 데이터로 구분하였다.

이에 비해 사회과학이나 인문학 분야 연구자들은 넓은 의미에서 관찰 데이터에 속하는 설문 데이터나 인터뷰 데이터를 주로 수집하고 있었으며 외부 기관에서 제공하는 데이터베이스 또는 정부통계 등의 기록을 데이터로 활용하고 있었다. 관광경영학 분야 연구자(P3)와 행정학 분야 연구자(P31)의 경우 거의 대부분 설문조사를 통해 데이터를 수집하고 있다고 응답하였다. 이들은 설문조사와 함께 인터뷰를 통해서도 데이터를 수집하고 있었으며 법학 분야 연구자

(P45), 정치외교학 분야 연구자(P51), 불교학 연구자(P16)의 경우도 인터뷰 데이터를 수집하는 경우가 빈번하였다. 기록을 활용하는 경우로 불교학 연구자(P16)나 종교학 연구자(P59)는 역사사료나 원전 문헌자료를 데이터로 인식하고 있었으며 법학 분야 연구자(P45)는 지적재산권과 관련된 민간기업의 근무규칙, 취업규정 등을 참고한다고 응답하였다. 이 연구자는 특허청이나 저작권위원회 등에서 발행하는 통계자료도 데이터로 활용하고 있었으며 관광경영학 분야 연구자(P3)의 경우 기존의 결과 값을 확인하는 용도로 정부통계를 이용하고 있다고 설명하였다.

이 외에도 외부 기관에서 제공하는 데이터를 사용하는 경영학 분야 연구자(P36)는 한국신용평가정보에서 제공하는 KisValue, 한국상장회사협의회에서 제공하는 TS2000, 미국 펜실베니아 대학 와튼 스쿨(Wharton School)에서 제공하는 와튼연구데이터서비스(Wharton Research Data Services, WRDS)를 주로 활용한다고 하였다. 이 연구자는 필요한 데이터를 수집하기 위해 직접 설문조사를 실시하지는 않으며 필요할 경우 금융감독원 또는 증권거래소 웹사이트를 통해 데이터를 수집하는 경우도 있다고 응답하였다.

3.2 데이터 기록화

데이터 기록화(data documentation)는 데이터의 생성방법 및 이유, 데이터의 의미, 구조와 내용, 변경 및 코딩에 관련된 정보를 제공하는 활동을 의미한다. 크게 두 가지 수준에서의 데이터 기록화가 필요한데 먼저 연구과제 수준

에서의 기록화로서 데이터가 생성된 연구과제, 데이터 생성과정, 데이터에 관련된 권한이나 일반적인 맥락에 대한 정보를 제공하는 것이다. 다음으로 데이터에 대한 기록화가 필요하며 이는 데이터의 내용, 구조, 형식에 대한 정보를 제공하는 것이다(Corti et al. 2014, 38).

특히 데이터에 대한 기술을 메타데이터 표준에 의거하여 제공하는 것은 효율적인 데이터 관리와 공유에 있어 필수적인 부분이라 할 수 있다. 공통된 메타데이터 표준을 활용함으로써 비슷한 데이터에 대한 기술의 일관성을 유지할 수 있고 데이터의 검색과 재이용을 촉진시킬 수 있다(Saltz et al. 2006; Karasti et al. 2010).

본 연구의 인터뷰 참여자들 중 데이터 기록화에 대하여 언급한 연구자는 소수에 불과하였으나 이들은 데이터에 대한 충분한 설명을 제공하는 것이 중요하다고 인식하고 있었다. 행정학 분야 연구자(P31)의 경우 IT 정책이 전공 분야이므로 메타데이터라는 용어를 인지하고 있다고 응답하였으며 데이터에 대한 메타데이터를 작성하려는 마음은 강하지만 여러 가지 이유로 제대로 작성하지 못하고 있다고 설명하였다. 그는 메타데이터를 “노트 형식”으로 작성한다고 하면서 “비고란에 적던지, 아니면 데이터가 좀 많은 경우에는 예를 들어서 녹취록[.....] 같은 경우에는 코딩을 하기 전에 엑셀에다가 자료들의 정리를 하면서 메타데이터”를 작성하는 방식을 취하고 있다고 하였다.

경영학 분야 연구자(P36) 역시 엑셀을 활용하여 메타데이터를 작성한다고 응답하였다. 그는 데이터 공유를 바람직한 현상으로 인식하고 있었으며 연구자 본인의 참고용으로 활용하거나 다른 연구자들이 데이터를 이해할 수 있도록

록 메타데이터를 제공하는 것이 필요하다고 보았다. 데이터 내의 변수명, 데이터가 생성된 과정 등을 비교적 소상히 작성하려고 하지만 나중에 보았을 때 본인이 작성한 것인데도 불구하고 어떤 내용인지를 이해하기 어려울 때도 있다고 하였다. 그는 “인프라”나 “방법론”을 가지고 데이터를 “체계적으로 관리를 하고” 싶고 “항상 해야 된다는 생각은 있지만, 지금 당장 해야 할 것이 있기 때문에” 어려움이 있다는 점을 언급하였다.

예를 들어서 변수이름은 어떤 거고 뭐 이런 것은 어떤 과정을 거쳐서 했다던가, 이를테면, 예를 하나 들어볼게요. 저희는 경영자에측공시 뭐 이런 게 있거든요. 경영자들이 미리 예측을 했다고 보는 건데, 이런 건 1년에 여러 번 해요. 그러면 뭐 최초의 것만, 첫 번째 한 것만 모았다던가. 그런 식의 이게 대체 어떤 과정을 통해서 이게 된 것이다. 그런 것을 알기 위한 것들을 가져다가 가능하면 충실히 하려고 해요. 가능하면 그냥 제3자가 보기에도 맞을 수 있도록 하려고 하는데도, 나중에 보면 아 이게 어떻게 된 거지? 그런 것들이 가끔씩 있습니다. (P36, 경영학, 조교수, 30대)

상기 경영학 연구자의 경우 다른 연구자와 데이터를 공유하는 상황을 염두에 두고 데이터에 대한 기술을 제공하는 것을 언급한 반면 컴퓨터공학 연구자(P47)의 경우 개발한 소스코드에 대한 설명을 제공하기는 하지만 이는 연구에 참여한 사람들이 이해하기 위한 목적으로 작성되는 것일 뿐 외부 연구자들은 그 내용을 이해하기 어려울 것이라고 응답하였다. 즉 컴퓨터 프로그램을 이해하기 전에 알아야 하는 기본

적인 지식과 경험이 상당히 많은데 이러한 능력 없이 연구에 참여한 사람들이 제공하는 설명만으로 소스 코드를 이해하기는 불가능하다는 것이다. 이러한 이유로 본인의 연구 분야에서는 소스코드를 다른 이들과 공유한다고 해도 제대로 활용될 수 없으며 이를 원하는 연구자들도 없을 것이라고 설명하였다.

3.3 데이터 저장 및 보존

인터뷰 참여자들은 일반적으로 데이터의 유용성이 유지되는 연한을 학술지 논문을 출판하는 활동과 연관해서 인식하고 있었다. 이러한 인식에 근거하여 몇몇 연구자들은 데이터 생성 또는 수집 후 1년에서 4년 정도를 데이터가 유용한 기한으로 응답하고 데이터가 최소한 그 기간만큼은 유지되어야 한다는 것에 동의하고 있었다. 경영학 분야 연구자(P36)의 경우 “너무 아웃데이트 되면 연구는 의미 없다고 생각하기 때문에” 1년에서 4년 정도의 기간이 데이터에서 논문을 출판할 수 있는 적절한 기한이라고 설명하였다. 그는 “주니어 교수들은 편수에 신경 써야 되다 보니까”, “옛날 자료를 모아서 축적해서” 연구하기보다는 “새로운 어떤 연구프로젝트가 있다든지, 아니면 새로운 연구 아이디어가 있으면 빨리 그거를 해야 되겠다”는 생각에 데이터 관리는 소홀하게 되는 경향이 있다고 하였다.

컴퓨터공학 분야 연구자(P47)도 데이터의 유용성은 4년 이내라고 응답하였고 그 이유로 무선 통신 영역에서 하나의 코드를 만들어놓으면 시간이 흐르면서 조금씩 코드를 바꾸어가며 사용할 수 있는데 3-4년이 지나면 기술이 충분

히 진화하면서 이전 코드를 더 이상 사용할 수 없게 되고 결국 새로운 코드를 개발해야 하는 상황이기 때문이라고 설명하였다. 분야는 다르지만 유사한 의견으로 법학 분야 연구자(P45)의 경우 본인의 연구 분야인 지적재산권 영역은 다른 법학 연구 영역에 비해 빠르게 변화하기 때문에 4년 이상 된 데이터를 논문에 활용할 수는 없다고 응답하였다.

위낙 이 분야에서의 발전 속도가 되게 빠르더라고 요. 그래서 사실 법학 관련 데이터의 수명이 그렇게 짧지는 않거든요, 다른 전공에는. 그런데 제 경우에는 법률이나 이런 것들도 너무 자주 바뀌는 분야이고, 시장 수요도 많이 바뀌고, 침해의 형태도 되게 많이 바뀌어서, 거의 한 1년에서 2-3년 지나면은 다시 한 번 그 분야를 업데이트 하거나 새로운 자료를 참조하지 않으면 한 4년 이상 된 자료를 가지고 최근 경향이 이렇다. 라고 쓸 수는 전혀 없는 거라서. (P45, 법학, 부교수, 40대)

데이터의 유용성을 5년 이상으로 보는 연구자들의 경우 데이터를 수집할 당시의 목적과는 다른 연구 관점에서 데이터를 활용하려는 경향이 있었다. 경영학 분야 연구자(P37)의 경우 어렵게 데이터를 수집하여 필요한 만큼의 논문을 출판하였지만 향후 비슷한 연구를 수행하여 비교 연구를 진행할 수도 있겠다는 생각이 들어서 데이터를 지속적으로 보관하고 있다고 응답하였다. 한의학 분야 연구자(P62)는 데이터 생성 이후 5년에서 10년 정도는 충분히 활용할 수 있을 것이라고 응답하였는데 한의학을 고문헌뿐만 아니라 실제 데이터에 의거하여 강의하는 사례를 제시하면서 교육 목적의 데이터 활

용 가능성을 언급하였다.

교육목적으로 주로 사용하고요. 그러니까 과거의 데이터를 대학원생들한테 강의를 할 때, 내가 이런 이런 연구를 해보니까 이런 이런 결과가 있더라. 제가 한약이란 것을 학생들한테 강의를 하기 때문에. 그 때 과거의 이야기, 본초강목 이런 이야기만 하면 안 되지 않습니까. 그 하단에 현대적인 연구 동향면에 제가 연구했던 것은 제 연구결과를 보여주고, 또 제 연구결과가 없으면 다른 연구결과를 보여주고 이런 식으로 수업을 진행합니다. (P62, 한의학, 조교수, 30대)

데이터의 유용성을 10년에서 19년 정도로 본다고 응답한 토목공학 연구자(P49)의 경우 본인이 수행하고 있는 지진공학 연구 동향이 그 정도 주기로 변화할 것이라고 예상하기 때문이라고 설명하였다. 데이터의 유용성을 무기한으로 보고 있다는 불교학 연구자(P16)는 본인이 문헌연구나 면담을 통해 분석한 데이터가 책이나 논문의 형태로 출간되고 그 출판된 자료에 대한 독자가 있는 한 그 유용성은 계속될 것이라고 응답하였다.

데이터가 유용하다고 인식하는 기간과는 별개로 연구자들은 생성 혹은 수집한 데이터를 지속적으로 보관하는 경향이 있었다. 경영학 분야 연구자(P37)의 경우 데이터의 양이 그렇게 크지 않기 때문에 사용했던 데이터와 연구 모형 등은 가능하면 보관을 하고 있다고 응답하였다. 박사학위 논문 데이터부터 본인이 주저자로서 출판하는데 활용했던 데이터는 생산된 지 6-7년 정도 되었지만 모두 가지고 있다고 하였다. 법학 분야 연구자(P45)의 경우도 연구

분야는 빠르게 변화하지만 데이터는 “언젠가 또 활용할 수 있는 거라는 생각”이 들고 “데이터의 변화과정을 추적하는 것도 최근 경향을 파악하는데 도움이 된다”는 생각에서 수집한 데이터를 가급적 폐기하지 않고 보존하고 있다고 설명하였다. 토목공학 연구자(P49)의 경우 본인의 데이터를 굳이 폐기할 생각은 없으며 “연구생활을 하는 동안” 약 “30년 정도” 보존할 의사를 가지고 있었다. 이 외에도 종교학 분야 연구자(P59)는 그 데이터를 수집했을 때 소요되었던 비용, 데이터로서의 가치, 그리고 본인의 연구 이력을 나타낸다는 상징적 의미에서 그간의 데이터와 자료들을 보관하고 있다고 설명하였다.

그거 뭐 내버리기 아깝고, 복사비만 해도 엄청나니까. 그리고 자료는 또 자료로 소중하고 나중에 혹시 가끔이라도 쓸모가 있을 수 있고, 그리고 추억의 가치도 있고, 제가 연구한 이력을 그래도 담고 있는 거니까. 그래서 보관을 해놓는 거죠 (P59, 종교학, 정교수, 40대)

이에 비해서 연구데이터의 폐기를 실행한다고 응답한 관광경영학 분야 연구자(P3)는 연구를 수행한지 몇 년이 지나면 해당 영역에 대한 연구를 더 이상 수행하지 않을 것이라는 생각이 들고 특별히 관심을 가지는 데이터가 아닌 한 폐기를 한다고 설명하였다. 주로 설문 데이터를 수집하는 이 연구자는 1년에서 4년 정도의 기한 이후 필요하지 않다고 판단되는 데이터는 컴퓨터 저장 공간에서 삭제를 한다고 응답하였다. 그는 논문에 제시된 결과 값을 가지고 데이터를 역으로 파악할 수도 있기 때문

에 데이터가 굳이 보존될 필요가 없다고 인식하고 있었다. 통계학 연구자(P41) 역시 컴퓨터 저장 공간 확보를 위해서 이미 오래된 방법론을 활용한 데이터는 폐기한다고 응답하였다. 그는 시뮬레이션 함수만 가지고 있으면 시뮬레이션을 통해 필요한 데이터가 생성이 되기 때문에 그 결과물인 데이터는 삭제한다고 설명하였다. 불교학 분야 연구자(P16)는 어렵게 구한 문헌이나 데이터일 경우를 제외하고 다른 방식으로 쉽게 구할 수 있는 자료일 경우는 폐기한다고 응답하였다. 예를 들어 구술을 받아 녹음한 자료일 경우 본인의 연구 분야에서는 드문 데이터이기 때문에 보존을 하지만, 구하기 힘든 문헌이라 복사를 해 놓았는데 정부에서 이를 데이터베이스화 했을 경우 그러한 문헌들은 폐기를 한다고 설명하였다.

3.4 데이터 공유

데이터를 공유하고 있다고 응답한 연구자들 대부분은 개인적인 연구 소모임이나 소속연구팀 내에서만 공유하고 있는 것으로 나타났다. 지질학 분야 연구자(P50)의 경우 같은 실험실 내에서도 공동 연구하는 팀과는 학술논문이 게재되기 이전에도 데이터를 공유하지만 그 이외의 사람들에게 특히 논문이 게재되기 이전에는 데이터를 비공개한다고 언급하였다. 경영학 분야 연구자(P37)의 경우 같은 지도교수님 아래 박사학위를 받은 현직 교수들의 연구 소모임이 있어 그들과 한 달에 한 번씩 학술세미나를 하면서 연구주제를 토론하고 공동연구를 수행하기도 한다고 설명하였다. 그는 이러한 소모임 안에서 “같이 공동 연구하는 연구자들하고는

데이터 공유하는 문제에 대해서는 제한을 두지” 않는다고 하였다.

법학 분야 연구자(P45) 역시 “공식적인 학회는 아니면서 [.....] 10명에서 20명 내외에서 연구자 모임들 같은 것들, 조찬모임 등을 통해서 공유하는 자료”가 존재한다고 설명하였다. 이 연구자가 속한 연구모임에도 학계 연구자뿐만 아니라 지적재산권 관련 변호사나 변리사, 산업계 종사자 등 다양한 관련 분야 종사자들이 포함되어 있었다. 연구모임에서 공유하는 데이터 중에는 “클라이언트들에 대해 관련된 부분”이나 “사건의 어떤 다른 배경에 관련된 부분” 또는 “고객의 어떤 영업상의 비밀 문제” 등이 포함되어 있는 경우가 많아 외부에 공개할 수 없고 연구모임 내부에서만 공유하게 된다고 설명하였다.

외부 연구자들이 요청할 경우에는 기꺼이 공유한다고 응답한 경영학 분야 연구자(P36)의 경우 개인적으로 친한 연구자일 경우 좀 더 편하게 공유할 수 있고 데이터에 대한 이해를 돕기 위해 간단한 예제도 만들어서 함께 제공한다고 하였다. 그는 전혀 모르는 연구자가 데이터를 요청을 하는 경우에도 공유를 하고 있다고 하였다. 그렇지만 본인의 연구 분야에서 데이터 공유는 전혀 활성화되어 있지 않고 심지어 “그런 이야기를 꺼내는 것조차도 실례”라는 분위기가 있기 때문에 데이터 공유를 드러내놓고 하는 것이 조심스럽다는 점을 언급하였다.

데이터를 누구나 볼 수 있게 공유해야 한다고 응답한 연구자들은 특히 국가에서 연구비를 지원받은 연구라면 반드시 그렇게 해야 한다고 주장하였다. 불교학 연구자(P16)는 한국연구재단에서 지원을 받아 생산한 구술 면담자료도

모두가 볼 수 있게 공유할 수 있다고 언급하면서 “개인 사비를 들여서 한 것도 아니고, 국가에서 해준 걸 들여서 한 것이니까 당연히 다 나와야 되는” 것이라고 설명하였다. 한의학 연구자(P62) 역시 정부기관에서 지원을 받아 생산된 데이터는 “국가에서 돈을 받은 것이기 때문에 당연히 모두의 것”이라고 인식하고 있었다. 종교학 연구자(P59)의 경우 국가에서 연구비를 지원받아 연구를 수행했다면 “국가적인 채무”를 가지게 되는 것이므로 “적어도 동료학자들에게는 공개”하고 “관심 있는 동료학자가 아닌 공동체 바깥에 있는 사람들한테도 공개하는 게 필요”하다고 생각하고 있었다. 또한 모든 연구자들이 연구를 수행하는 과정에서 “동학들이나 제자들”에게 음으로 양으로 도움을 받게 되므로 이는 “사회적인 채무”를 갖게 되는 것이라고 설명하였다. 그는 모든 연구자들이 국가와 사회, 동료 연구자들에게 “배움의 빛”이 있기 때문에 “빛진 자 의식”을 가지고 연구의 결과를 독점하기보다는 공개하는 것이 바람직하다고 주장하였다.

그러나 국가에서 지원받은 연구의 결과물은 공공재이므로 공개해야 한다는 논리에 부분적으로 동의하거나 동의하지 않는다는 연구자들도 있었다. 행정학 분야 연구자(P31)는 미국의 예를 들면서 미국인들은 세금에 대한 개념이 확고하기 때문에 저명한 국립과학재단 등지에서 연구비를 받는 것을 굉장히 영예롭게 생각하고 그 결과물을 공개하는 것에 전혀 이의가 없지만 이러한 논리를 국내에 그대로 적용하는 것은 무리가 있을 것이라고 보았다. 경영학 분야 연구자(P37)의 경우 부분적으로는 동의하지만 연구비를 원하는 만큼 충분히 지원받지 못

하는 상황에서 단순히 국가로부터 지원받았다는 이유로 연구 결과물을 무조건 공유해야 한다는 것은 모순이 있다고 설명하였다.

뒤 100% 동감하지 않지만, 어느 정도는 동감을 합니다. 한 70~80%는 동감을 합니다. 다만 그 연구비라는 게 사실 사회에서 지원되는 게, 또 어떻게 보면 연구자가 원하는 만큼 주는 것이 아니잖아요, 어떤 정해진 제약 속에서 제공을 하는 거고[.....] 분명히 그 가치만큼은 어떻게 보면 공개를 하는 것에 대해서는 찬성을 하지만, 또 일부 연구비 이상으로 연구된 어떤 공헌이나 아니면 그런 정성이 들어가 있는 부분이 충분히 있을 거라고 판단이 되기 때문에 무조건 이게 국가로부터 지원을 받았다고 해가지고 그 정보도 무조건적으로 공유된다 이것은 조금 논리적으로 안 맞는 것 같아요. (P37, 경영학, 조교수, 30대)

관광경영학 분야 연구자(P3)의 경우 국가연구비를 지원받았다고 해도 데이터를 공유하는 것은 반대하는 입장을 밝혔다. 그는 학술지 논문을 출판하는 과정을 통해 이미 데이터의 분석 결과가 공개된 상태임을 강조하였다. 따라서 연구비를 지원하는 기관에서 데이터를 공유하라고 한다면 그것은 “충분한 설명력이 있어야 가능한 것이지 일방적으로 데이터를 달라고 하는 것”으로 되어서는 안 된다고 주장하였다. 그는 연구자들이 너무 쉽게 다른 사람들의 데이터를 이용해서 논문을 출판할 수도 있고 연구 아이디어가 노출되어 표절의 위험성이 있다는 점에서 데이터 공유에 부정적인 견해를 보였다.

법학 분야 연구자(P45) 역시 연구데이터 공

유에 있어서 표절이나 아이디어 도용, 지적재산권 문제를 우려하고 있었고 그렇기 때문에 “연구 도중에 데이터가 공유되는 것은 조금 꺼려지게” 된다고 설명하였다. 경영학 분야 연구자(P37)의 경우 이와 관련하여 연구 소모임에서 향후 연구계획에 대한 아이디어를 공유했는데 모임에 참여하는 선후배들이 그 주제에 대한 연구를 앞 다투어 진행하게 되면서 부득이 본인의 연구 주제를 바꿀 수밖에 없었다는 이야기를 하였다. 그는 연구 영역을 바꾸게 된 이유로 연구 소모임에 있는 연구자들이 같은 분야 동료이면서 선후배이기 때문에 “기본적인 관계를 깨뜨리고 싶지” 않았기 때문이라고 설명하였다.

지질학 분야 연구자(P50)는 연구의 독창성과 관련해서 본인의 연구 분야에서는 학술지 논문이 출판됨으로써 누가 처음으로 그 연구를 발표했는지가 결정이 된다고 하였다. 따라서 데이터가 논문 출판 이전에 공유된다면 그 데이터를 가지고 논문을 출판하는 사람이 주도권을 가지게 되는 것이므로 논문이 출판되기 이전에는 데이터의 공유를 엄격히 제한한다고 설명하였다.

그걸 공개하면, 논문으로 나왔을 때 originality가 없거든요. 저희는 그게 중요하거든요. 누가 먼저 이것을 발표를 했는지가 통계자료거나 그런 게 아니라 순수 그 암석에서, 아니면 저희 샘플에서 얻어진 데이터이기 때문에 [데이터]는 공개가 되려면 정식 게재된 후에. (P50, 지질학, 박사후연구원, 20대)

토목공학 연구자(P49)의 경우 본인의 연구

분야에서 수행한 연구의 결과는 “이미 절차가 밝혀진 해석 결과”이기 때문에 “결과데이터를 공유한다고 해서 아이디어가 도용될 우려는 없다”고 설명하였다. 그러나 결과를 해석하면서 오류를 범할 가능성이 있고 데이터를 공유하지 않았으면 발견되지 않았을 오류를 공개하게 될 우려가 있다는 점을 지적하였다.

제가 해석하면서 뭔가 오류를 범했을 경우에, 그리고 제가 그걸 가지고 제가 연구 결과를 발표했는데 나중에 다른 분이 그 데이터를 가지고 봤더니 뭐가 잘못되었더라 라고 했을 때[.....] 공유를 안 하면 오류가 밝혀질 확률이 0%잖아요, 근데 공유를 하면 확률이 생기니까. 그래서 좀 꺼려하는 것 같아요. (P49, 토목공학, 부교수, 40대)

또한 상기 연구자는 데이터를 공유할 때 “그냥 raw data만 주고서 다른 사람이 쓸 수 있는 건” 없기 때문에 “의미 있는 결과로 분석하는 절차”를 제공해 주어야 하는데 그 작업에는 추가적인 시간과 노력이 필요하다고 설명하였다. 그러나 다른 업무가 많은 입장에서 그러한 노력을 들이기가 사실상 어렵다고 하였고 “나중에 여유가 있으면” 데이터 공유에 참여할 의사가 있다고 언급하였다.

이와 같이 인터뷰 참여자들은 데이터를 공유한다고 해도 개인적인 연구소모임이나 연구팀 내에서만 공유하는 경향이 있었으며 외부 연구자의 요청이 있으면 공유를 하거나 모두가 볼 수 있게 공유한다는 의견은 소수에 불과하였다. 특히 다수의 해외 정부연구비지원기관에서 천명하는 것처럼 국가연구비 지원을 받은 과제의 데이터는 공공재이므로 공유해야 한다는 논리

에 일부 연구자들은 전적으로 동의하였으나 부분적으로 동의하거나 동의하지 않는 경우도 있었다. 데이터 공유에 있어서 연구자들은 표절이나 아이디어 도용에 대한 우려, 논문 출판의 주도권을 놓칠 수 있다는 위험, 혹시나 존재할 수도 있는 오류가 노출될 수 있다는 우려, 데이터 공유에 소요되는 시간과 노력에 대한 부담감을 가지고 있는 것으로 나타났다.

3.5 데이터 재이용

다른 연구자들의 데이터를 재이용한 경험이 있는 인터뷰 참여자들 중 대부분은 논문에서 데이터를 추출하여 사용하거나 추가적인 데이터를 얻기 위해 논문 저자에게 문의하는 방식으로 타인의 데이터를 획득하고 있었다. 토목공학 연구자(P49)의 경우 컴퓨터로 해석한 본인의 연구결과를 실험 연구를 수행한 다른 연구자의 연구 결과와 비교 분석하는 연구를 수행할 때 그 연구자의 논문에 제시되어 있는 실험 시편의 정보를 가져다 본인의 해석에 활용을 한다고 설명하였다. 컴퓨터공학 연구자(P47)는 다른 연구자의 논문에 제시되어 있는 모의실험 환경을 충실히 재생산해 내고 그 결과도 기존의 논문에 나와 있는 것과 유사하게 나왔다는 것을 보여주면서 연구자 본인이 새로운 알고리즘 요소를 추가하여 더 좋은 결과를 도출하였다면 제시한 알고리즘에 대한 학계의 신뢰를 얻을 수 있다고 설명하였다. 이와 같이 기존 알고리즘의 결과를 증명하고 새롭게 개선하는 연구를 수행하는데 있어 다른 연구자들의 논문에 제시된 모의실험 환경에 대한 데이터를 활용할 수 있음을 언급하였다. 그는 본인의 연구 분야에서 다른 연구자

들이 개발한 알고리즘 코드를 데이터로 활용하기 위해 요청하는 것은 “굉장히 무례한” 일이지만 논문에 제시된 데이터는 공개된 것이므로 활용하고 있다고 언급하였다.

다른 연구자들의 데이터를 활용하기 위해 논문을 먼저 찾아보지만 실험에 관련된 더 자세한 내용을 알고 싶어 개인적으로 연구자들에게 연락을 해서 원하는 데이터를 얻는다는 연구자도 있었다. 한의학 분야 연구자(P62)의 경우 논문에 “표시되지 않는 진짜 노하우들이 있는데” 출판된 논문만 보고 실험을 기획한다는 것은 상당히 어려운 일이기 때문에 연구자들에게 직접 연락하여 실험의 세부사항을 문의한다고 하였다. 그는 전혀 모르는 연구자에게 연락을 해 본 적은 없고, 개인적으로 아는 연구자이거나 아는 사람을 통해서 연락이 가능한 연구자들에게만 문의를 했다고 설명하였다. 문의를 했을 경우에는 누구나 흔쾌히 실험 방법을 알려주었고, 실험실에 방문하게 해서 직접 실험의 과정을 보여주는 등 실험의 노하우를 100% 공개하는 연구자도 있었다고 하였다.

이 외에도 불교학 연구자(P16)의 경우 다른 연구자들의 책이나 논문을 보면서 데이터의 출처를 확인하고자 할 때 직접 저자들에게 연락을 취한다고 하였다. 본인의 연구 분야가 불교 철학이기 때문에 사료 등 원 자료를 다루는 것이 쉽지 않기 때문에 원 자료에 대해 많은 지식을 가지고 있는 스님이나 역사학자들에게 주로 문의를 한다고 설명하였다.

어떤 분이 책이나 논문을 냈을 때, 제가 이렇게 보고, 이거는 뭘 보고 확인하셨을까 싶은 생각이 드는 부분이 있어요, 그러면 제가 직접 전화를

하거나 메일을 보내가지고 요거는 어떻게 확인을 하느냐, 어떤 자료를 보셨느냐 이렇게 제가 다시 여쭙죠. [.....] 주로 저는 여쭙는 분들이 주로 스님들이나, 또 아무래도 사료를 많이 다루는 분들은 사학 쪽을 하시는 분들이니까 역사학 쪽. [.....] 저는 철학분야니까 사료를 직접 다루기는 제 능력 밖인 경우가 많아요. 예를 들면 초서도 봐야 하고, 또 다른 고전문헌들을 아무래도 [누군가] 가지고 있고, 전문적으로 다루는 분들은 역사학 쪽의 분들이니까요. (P16, 불교학, 조교수, 40대)

이와 같은 방법 이외에 직접 다른 연구자들의 데이터를 활용한다는 연구자들은 소수에 불과하였다. 경영학 분야 연구자(P37)는 다른 연구자들에게 데이터를 요청하는 경우 “그 주제와 절대적으로 관련이 있는 데이터만 요청”을 한다고 하였다. 그러한 이유로 “아직까지는 공유를 한다는 분위기가 안 된 상태이기 때문에” 모든 데이터를 요청하는 것은 오해의 소지가 있으므로 조심스럽게 접근할 수밖에 없다고 설명하였다. 통계학 연구자(P41)의 경우 논문을 통해 데이터의 출처가 명시가 되어 있으면 그 출처를 보고 데이터를 얻는 것이 일반적이라고 하였다. 그렇지 않은 경우 데이터를 가지고 있는 사람의 정보가 주어지기도 하는데 개인적으로 연락하더라도 데이터를 공유하지 않는 경우가 상당수 있었다고 언급하였다. 그렇지만 외국의 데이터는 많이 공개되어 있기 때문에 구글 검색을 통해서 비교적 쉽게 구할 수 있다고 설명하였다.

외국은 데이터 사이트가 잘 정리가 되어있어요,

퍼블릭 데이터들은, 그리고 오픈 액세스가 많기 때문에 구글링하면 굉장히 쉽게 구할 수 있어요, 퍼블릭화되었지만, 구할 수 없는 자료는 직접 author에게 컨택을 할 수도 있구요.[.....] 그쪽 [에서] 쥐고 있는 데이터를 갖고 있는 사람이 내놓지 않는 이상은 어렵지 않게 데이터를 구하는 것 같아요. (P41, 통계학, 조교수, 30대)

인터뷰 참여자 중 일부는 외국과 비교하여 국내의 데이터 공유의 풍토가 미약한 수준임을 어렵게 생각한다는 의견을 제시하였다. 경영학 분야 연구자(P36)는 특정 분야에서 대가로 인정받는 외국 교수님들이 데이터를 홈페이지에 누구나 볼 수 있게 공유하는 사례를 예로 들면서 “데이터를 같이 나눠서 쓰고”, “모두가 공유를 하게 되면 [사회적 비용]이 훨씬 더” 감소할 수 있을 것이라 보았다. 한의학 분야 연구자(P62)는 데이터 공유를 촉진하는 학계와 사회의 분위기가 중요하다는 것을 강조하면서 본인의 연구 분야에서 정한용법이라는 침 법을 개발한 연구자는 동영상만 만들어서 모든 것을 공개한 반면 초기에 이를 배웠던 다른 사람은 1인당 돈을 받고 그 방법을 전수를 했다는 예를 제시하였다. 그는 어느 분이 존경을 받을지는 답이 나오는 것이지만, “어떤 지식을 공개했을 때 그것이 존중받고 존경받는 사회 분위기가 중요”하며 그러한 풍토가 국내에도 조성되는 것을 희망한다는 의견을 보였다.

4. 요약 및 제언

본 연구에서는 대학 연구자 13명의 인터뷰를

통해 연구데이터의 관리와 공유에 대한 인식을 조사하였다. 아직 국내에서는 연구데이터 관리 서비스에 대한 논의가 활성화되어 있지 않은 상황이지만 데이터 큐레이션 등 데이터를 보존하고 지속적인 서비스를 제공하는 해외 동향에 대한 관심이 높아지고 있는 실정이다. 본 연구의 결과를 통해 데이터 관리서비스를 위한 시사점을 제시하면 다음과 같다.

첫째, 연구자들이 다루고 있는 데이터의 유형으로 이공 및 의학 분야 연구자들은 실험데이터 또는 계산데이터를 주로 생산한다고 응답하였다. 특히 계산데이터의 경우 시뮬레이션 함수나 알고리즘이 중요한 가치를 가지는 데이터로 인식되고 있었다. 이를 통해 어떠한 모의 실험을 통해 생성되는 결과 값보다 그 모의 실험을 가능하게 하는 함수와 알고리즘이 연구의 핵심이라는 것을 알 수 있다. 그러나 물리적 혹은 화학적 실험을 통해 생성되는 데이터의 경우 그 결과 값으로 나타나는 데이터가 중요한 의미를 가지게 된다. 이러한 실험데이터는 다양한 실험 장비를 통해 생산되거나 실험에 참여한 연구자의 육안으로 관찰된 내용을 기록하는 방식으로 생성된다. 이에 비해 사회과학이나 인문학 연구자들의 데이터는 주로 설문이나 인터뷰를 통해 생성되거나 외부 기관에서 생성되는 원 자료, 정부통계 또는 상업적인 데이터베이스, 웹사이트, 역사적인 문헌에 이르기까지 다양한 유형으로 제시되었다.

인터뷰 참여자들이 언급한 데이터는 대부분 디지털 형태의 데이터였으나 인문학 연구자들의 경우 물리적인 형태로만 획득할 수 있는 자료나 문헌자료는 종이 복사본의 형태로 보관하고 있다고 응답하였다. 인터뷰에서 언급된 디

지털 데이터의 포맷은 엑셀과 같은 스프레드시트, 동영상 데이터, 면담 녹취 파일과 같은 오디오 데이터, 설문지나 면담녹취록 또는 알고리즘 코드 파일 등의 문서류, 암석에서 채취한 샘플, 데이터베이스, 웹사이트 등 다양하였다. 데이터의 양에 있어서도 컴퓨터 저장 공간을 확보하기 위해 필요 없는 데이터는 수시로 삭제한다는 연구자들이 있는 반면 데이터가 저장 공간을 그리 많이 차지하지는 않기 때문에 최근 6-7년 동안의 데이터를 모두 보관하고 있는 연구자도 있었다.

이러한 연구 결과를 통해 효율적인 연구데이터 관리서비스를 위해서는 다양한 데이터의 유형과 포맷에 대한 실무자들의 이해와 지식이 요청된다는 것을 알 수 있다. 연구 분야 별로 중요하게 인식되는 데이터의 종류와 데이터 양의 증가 속도에도 차이가 있으므로 이러한 차이를 염두에 두어 관리 대상이 되는 데이터의 유형과 종류를 정하는 것이 필요하다. 또한 향후 데이터의 보존을 고려할 때 보존에 적합한 개방형 포맷을 중심으로 각 유형별 데이터 포맷에 대한 정책이 마련될 필요가 있다.

둘째, 데이터 기록화에 있어 본 연구에서는 소수의 참여자들만이 데이터 생성이나 수집의 맥락에 관한 설명을 정리하는 작업을 간단하게나마 수행하고 있었다. 실제로 데이터를 생산하는 연구자가 연구의 맥락과 데이터 생성방법, 데이터의 특성에 대한 충분한 설명을 제공하는 것은 데이터의 관리와 공유에 있어 필수적인 작업이라고 할 수 있다. Lynch(2008)에 따르면 데이터 기술과 교환을 위해서 연구 분야별로 데이터에 대한 메타데이터 표준을 정의하고 연구자들이 이를 적용하는 사례가 증가하고 있다고 하였다.

생명과학 분야의 Darwin Core나 생태학 분야의 Ecological Metadata Language(EML) 등은 그러한 시도의 대표적인 예라고 할 수 있다.

인터뷰 참여자들 중에는 메타데이터의 개념을 인지하는 소수의 연구자들이 있었으며 한 연구자는 데이터를 기술하는 인프로나 방법론 등 데이터의 체계적인 관리를 가능하게 하는 메커니즘에 대한 요구를 가지고 있었다. 또한 데이터의 관리나 기술을 수행하는 작업 자체가 다른 업무에 비해 우선순위가 낮다보니 결국 미루게 된다는 의견에서도 알 수 있듯이 데이터의 기록화를 지원하는 기능이 연구데이터 관리서비스의 중요한 요소로 제공되어야 할 것이다. 정책적인 측면에서 데이터 기록화 및 메타데이터의 개념과 그 필요성을 연구자들에게 인식시킬 필요가 있으며 각 연구 커뮤니티별로 정의되거나 실제 활용되고 있는 메타데이터 표준이 있는지를 조사하여 이를 연구자들에게 안내하는 것이 필요할 것이다. 미국 조지아텍(Georgia Tech) 기관 레포지터리의 경우 데이터에 대한 기본적인 메타데이터 요소를 입력할 수 있는 템플릿을 제공하는데 이러한 템플릿을 활용한다면 최소한의 데이터 기록화를 유도할 수 있을 것이다 (Georgia Tech Library 2013).

셋째, 인터뷰 참여자들이 인식하는 데이터 유용성의 기간은 논문에 데이터를 활용할 수 있는지의 여부와 밀접하게 관련된다는 것을 알 수 있었다. 다수의 참여자들이 1년에서 4년 정도를 데이터가 유용한 기간이라고 인식하고 있었으며 그 이후에 해당 데이터를 가지고 논문을 출판하기는 어렵다고 하였다. 그러나 연구자들은 데이터가 유용할 것이라고 인식하는 기간보다 더 오랜 기간 데이터를 보관하고 있는

것으로 나타났다. 이에 대해서 향후 데이터를 비교 연구 등 다른 방법으로 활용할 수 있을 것이라는 생각에서 유지하는 경우와, 강의 등 교육적인 목적으로 데이터를 활용할 수 있기 때문이라는 의견도 있었다. 그러나 데이터가 연구동향을 반영하는 시기가 10년에서 20년 사이로 상대적으로 긴 경우도 있었으며 인문학 연구자들은 본인이 저술한 책이나 논문에 대한 독자가 존재하는 한 그 자료의 기반이 되는 데이터의 유용성도 무기한일 것이라고 응답하기도 하였다.

이와 같이 연구자들이 실제로 데이터를 보존하는 기간은 적어도 5년 이상인 것으로 나타났으며 컴퓨터의 저장 공간이 허용하는 한 폐기하지 않고 전부 보관하는 경우가 많았다. 폐기한다는 이유의 대부분은 저장 공간을 확보하기 위해서였으며 이 경우 시물레이션을 통해 생성할 수 있는 결과 값 데이터는 삭제한다고 하였다. 한 인문학 연구자는 본인의 연구 역사를 담고 있는 추억의 자료이기 때문에 쉽게 버리지 못한다는 응답을 하였으며 그 자료를 복사할 때 소요된 비용도 상당하였고 자료 자체로서의 가치도 있으므로 폐기하지 않고 보존한다고 설명하였다.

이러한 결과는 데이터의 저장이나 보존에 대한 연구자들의 요구가 적지 않다는 것을 보여준다. 기존의 선행연구에서 저장 공간과 백업 설비의 확대 등은 연구데이터 관리서비스를 제공하는 초기에 연구자들의 참여를 촉진할 수 있는 요소가 될 수 있다고 하였다(Parsons et al. 2013). 그러나 데이터 관리서비스에서의 보존은 활용을 근간으로 하는 개념이므로 개인적인 목적의 저장과 보존만을 염두에 두고 있는 연구자일

경우 외부 기관이 주체가 되는 서비스에 데이터를 제공하지 않을 수도 있다. 또한 데이터의 보존활동은 단순한 백업이나 저장 공간의 제공이 아니라 보존을 지속적으로 지원하는 정책과 자원 및 기술적 인프라를 기반으로 수립되어야 할 것이다.

넷째, 데이터의 공유와 재이용은 대부분 개인적으로 신뢰하는 사람들 간에 제한적인 범위로 이루어지고 있음을 확인할 수 있었다. 이는 Wallis et al.(2013)가 제시한 대로 신뢰를 바탕으로 하는 학술 커뮤니티에서의 나눔의 문화(gift of culture in scholarship)를 반영하는 것이라고 할 수 있다. 하지만 일부 연구자들은 국가연구비로 지원받은 과제에서 생성된 데이터만큼은 누구나 볼 수 있게 공개해야 한다는 의견을 나타냈다. 그러나 이러한 논리에 부분적으로 찬성하거나 반대하는 의견도 있어 해외의 연구비 지원기관에서 오픈액세스 공유의 기본 원칙으로 제시되는 내용이 국내에서는 전적으로 받아들여지지 않는다는 것을 알 수 있다. 연구자들은 연구아이디어의 도용, 표절 등의 문제, 데이터의 오류가 발견될 가능성, 데이터를 공유하는데 드는 시간과 비용의 문제, 논문 출판의 주도권을 빼앗길 수 있다는 점 등을 우려하고 있는 것으로 나타났다. 데이터의 재이용에 있어서도 논문에 제시되어 있는 데이터를 활용한다는 의견이 많았으며 논문에 제시되어 있는 것 이상을 원할 경우 직접 논문 저자에게 문의하여 데이터를 얻는다는 연구자들도 있었다. 실제 다른 연구자들에게 데이터를 요청하는 경우에도 꼭 필요한 부분에 대한 것만 요청하여 오해의 소지가 없게 조심한다는 연구자들과 구글 검색을 통해 얻을 수 있는 공개된 외국 데

이터를 주로 활용한다는 연구자가 있었다.

이와 같이 국내에서 데이터 공유와 재이용은 제한적인 범위 내에서 이루어지고 있으므로 국제적인 추세인 데이터의 오픈 액세스를 추진한다면 다양한 유인책을 마련할 필요가 있을 것이다. 정부연구비로 지원받은 과제의 데이터를 공개하는 부분에 있어서 동의하는 연구자들이 상당수 있었으나 이에 완전히 동의하지 않는 연구자들도 있으므로 연구자들이 우려하는 바를 경감할 수 있는 정책적 요소가 필요하다. 예를 들어 영국의 정부연구비 지원기관인 웰컴 트러스트(Wellcome Trust)는 연구자들에게 독점적으로 데이터를 활용할 수 있는 기간을 제공하여 데이터 공유를 촉진하려는 방안을 제시하고 있으며 이는 논문 출판의 주도권을 빼앗길까 염려하는 연구자들에게 유인책이 될 수 있다(Wellcome Trust 2014). 미국의 국립과학재단(NSF)에서도 연구자가 생성 또는 수집한 데이터 및 연구결과물에 대한 모든 법적 권한을 인정해주는 정책을 제시하고 있고 이러한 정책은 표절이나 연구 아이디어 도용을 우려하는 연구자들의 참여를 유도할 수 있는 방안으로 활용될 수 있다(NSF 2013).

5. 결론

본 연구에서는 인터뷰데이터 분석을 통해 데이터의 생성 및 수집, 기록화, 저장 및 보존, 공유와 재이용에 대한 대학 연구자들의 다양한

경험과 의견을 조사하였다. 인터뷰 참여자 중 소수이기는 하지만 방법론이나 기술적인 인프라 등 체계적인 데이터 관리를 원하는 연구자들이 있었으며 국내에서도 해외의 선행연구에서 제시되고 있는 연구데이터 관리서비스의 정책적, 기술적 측면과 모범실무를 바탕으로 국내의 상황에 맞는 서비스 정책이나 모형에 대한 연구가 필요할 것이다. 특히 선행연구에서 제시된 대로 연구자들은 대학도서관을 데이터 관리의 주체로 인식하는 경향은 낮았으나 그 가능성을 긍정적으로 평가하고 있었으므로 대학에서의 연구데이터 관리서비스에서 대학도서관의 역할을 강화할 수 있는 방안을 논의할 필요가 있다.

아울러 데이터 공유와 재이용은 제한적인 범위 내에서만 이루어지고 있었으며 오픈 액세스 데이터 공유에 대한 연구자들의 의견 역시 다양한 것으로 나타났다. 연구자들은 데이터 공유의 근본적인 취지에는 공감하지만 시간과 노력 등 데이터 공유에 드는 비용과 공유로 인한 잠재적인 위험에 대한 우려를 표시하였다. 국내에서는 아직까지 정부나 연구 커뮤니티 차원에서 데이터 공개 및 공유를 적극적으로 지원하는 동력이 부족하지만 데이터가 전 세계적으로 학문 및 경제 발전의 자산으로 인식되는 상황에서 데이터 공유의 움직임은 더욱 활발히 전개될 것으로 예상된다. 따라서 연구자들의 우려를 최소화하면서 데이터를 공유하고 이를 활용할 수 있게 하는 정책과 기술에 대한 연구가 지속적으로 이루어져야 할 것이다.

참 고 문 헌

- [1] 김문정, 김성희. 2015. 과학기술분야 연구자의 연구데이터 공유의 영향요인에 대한 연구. 『한국문헌정보학회지』, 49(2): 313-334.
- [2] 김은정, 남태우. 2012. 연구데이터 수집에 영향을 미치는 요인 분석. 『정보관리학회지』, 29(2): 27-44.
- [3] Akers, K. G. and Doty, J. 2013. "Disciplinary differences in faculty research data management practices and perspectives." *International Journal of Digital Curation*, 8(2): 5-26.
- [4] Alexogiannopoulos, E., McKenney, S. and Pickton, M. 2010. *Research Data Management Project: a DAF investigation of research data management practices at The University of Northampton*. Northampton: The University of Northampton. [online] [cited 2015. 7. 15.] <<http://nectar.northampton.ac.uk/2736/1/Alexogiannopoulos20102736.pdf>>
- [5] Association of Research Libraries. 2014. *Open Scholarship*. Washington D. C., US: Association of Research Libraries. [online] [cited 2015. 7. 15.] <<http://www.arl.org/focus-areas/open-scholarship#.VaXCyKNWFMt>>
- [6] Borgman, C. L. 2012. "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology*, 63(6): 1059-1078.
- [7] Borgman, C. L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- [8] Borgman, C. L., Wallis, J. C. and Enyedy, N. 2007. "Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries." *International Journal on Digital Libraries*, 7(1-2): 17-30.
- [9] Corti, L. et al. 2014. *Managing and sharing research data: A guide to good practice*. London: SAGE Publications.
- [10] Cragin, M. H. et al. 2010. "Data sharing, small science and institutional repositories." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1926): 4023-4038.
- [11] Ekmekcioglu, C. and Rice, R. 2009. *Edinburgh Data Audit Implementation Project*. Scotland: The University of Edinburgh. [online] [cited 2015. 7. 15.] <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.175.6624&rep=rep1&type=pdf>>
- [12] Georgia Tech Library. 2013. *Cataloging Metadata Template*. Atalanta: Georgia Tech Library. [online] [cited 2015. 7. 15.]

- <<http://d7.library.gatech.edu/research-data/metadata>>
- [13] Goben, A. and Salo, D. 2013. "Federal research: Data requirements set to change." *College and Research Libraries News*, 74(8): 421-425. [online] [cited 2015. 7. 15.]
<<http://crln.acrl.org/content/74/8/421.full#ref-17>>
- [14] Karasti, H., Baker, K. S. and Millerand, F. 2010. "Infrastructure Time: Long-term Matters in Collaborative Development." *Computer Supported Cooperative Work-the Journal of Collaborative Computing*, 19(3-4): 377-415.
- [15] Kim, Y. and Stanton, J. M. 2012. "Institutional and individual influences on scientists' data sharing practices." *Journal of Computational Science Education*, 3(1): 47-56.
- [16] Lynch, C. 2008. "Big data: How do your data grow?" *Nature*, 455(7209): 28-29.
- [17] Molloy, J. C. 2011. "The open knowledge foundation: open data means better science." *PLoS Biology*, 9(12). [online] [cited 2015. 7. 15.]
<<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195>>
- [18] National Science Foundation, 2013. *Chapter VI - Other Post Award Requirements and Considerations*. National Science Foundation, Ann Arbor: National Science Foundation. [online] [cited 2015. 7. 15.]
<http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4>
- [19] Open Exeter Project Team, 2012. *Summary Findings of the Open Exeter Data Asset Framework (DAF) Survey*. Exeter: The University of Exeter. [online] [cited 2015. 7. 15.]
<<http://hdl.handle.net/10036/3689>>
- [20] Parsons, T., Grimshaw, S. and Williamson, L. 2013. *Research data management survey: report*. Nottingham: The University of Nottingham. [online] [cited 2015. 7. 15.]
<<http://eprints.nottingham.ac.uk/1893/>>
- [21] Saltz, J. et al. 2006. "caGrid: Design and implementation of the core architecture of the cancer biomedical informatics grid." *Bioinformatics*, 22(15): 1910-1916.
- [22] Scaramozzino, J. M., Ramírez, M. L. and McGaughey, K. J. 2011. "A study of faculty data curation behaviors and attitudes at a teaching-centered university." *College & Research Libraries*, 73(4): 349-365.
- [23] Wallis, J. C., Rolando, E. and Borgman, C. L. 2013. "If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology." *PloS one*, 8(7). [online] [cited 2015. 7. 15.]
<<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332>>
- [24] Wellcome Trust, 2014. *Establishing Incentives and Changing Cultures to Support Data Access*.

London: Wellcome Trust. [online] [cited 2015. 7. 15.]

<http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf>

[26] Westra, B. 2010. "Data services for the sciences: A needs assessment." *Ariadne*, 64. [online] [cited 2015. 7. 15.] <<http://www.ariadne.ac.uk/issue64/westra>>

[27] Whyte, A. and Pryor, G. 2011. "Open science in practice: Researcher perspectives and participation." *International Journal of Digital Curation*, 6(1): 199-213.

[28] Williams, S. C. 2013. "Data sharing interviews with crop sciences faculty: why they share data and how the library can help." *Issues in Science and Technology Librarianship*, 72. [online] [cited 2015. 7. 15.]

<http://www.istl.org/13-spring/refereed2.html?utm_source=dlvr.it&utm_medium=twitter>

• 국문 참고자료의 영어 표기

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

[1] Kim, Moonjeong and 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.

[2] Kim, Eun-Jeong and Nam, Tae-Woo. 2012. "Factor Analysis of Effects on Research Data Collection." *Journal of the Korean Society for Information Management*, 29(2): 27-44.