

연구데이터와 전문도서관

김선태 _ 전북대학교

마이크로소프트의 회장인 빌게이츠는 4세대 연구패러다임을 주장한 짐 그레이(Jim Gray) 생각에 대해 다음과 같이 주장했다. “데이터와 소프트웨어가 과학을 한다고 재정의 함으로써, 과학에 대해 우리가 새로운 방법으로 사고하도록 했다”. 이것은 연구 데이터의 중요성을 역설한 것으로 볼 수 있다.

해외 주요 선진국에서는 이러한 연구 데이터의 수집 및 저장, 관리, 보존, 출판, 서비스를 위한 다양한 활동을 하고 있으며, 이와 관련된 전문 인력을 채용하고 있다. 국내에서도 체계적인 연구 데이터 관리와 재사용을 보장하기 위해, 연구데이터 통합 관리 체계를 구축 중에 있다. 본고에서는 연구 데이터와 관련된 국내외 주요 동향을 살펴보고 도서관 현장에서 무엇을 준비해야 하는지 알아본다.

기하급수적으로 생산되는 연구 데이터



〈그림 1〉 체스와 기하급수

다음은 많은 체스 유래 중 하나다. 인도에 시네그람이라는 호전적인 왕이 있었다. 전쟁을 좋아하는 까닭에 무고한 백성이 전장에 끌려가 목숨을 잃었다. 그 나라에 세타라는 승려가 있었다. 세타는 체스를 만들어 왕에게 주었다. 그 후, 왕은 체스에 빠졌고 백성의 삶은 편안해졌다. 시네그람 왕은 세타를 궁으로 불러들였고 소원을 물었다. 세타는 체스판 첫 번째 칸에 밀 한 톨, 두 번째 칸에 밀 두 톨, 세 번째 칸에 밀 네 톨, 네 번째 칸에 밀 열여섯 톨, 그렇게 배수에 배

수로 체스판을 밀알로 채워 달라 말했다. 기껏 소원이 그것이나며 왕은 신하들에게 세타의 소원대로 밀을 채워주라 했다. 하지만 체스판 중간 쪽에 달했을 때, 당시 인도에서 생산되던 밀로도 충당하기 어려웠다. 결국 시네그람 왕은 세타를 옥심꾸러기라며 죽이고 만다. 이렇게 체스 전설을 장황하게 늘어놓는 이유는 '기하급수'를 설명하기 위함이다. 우리는 일상에서 급속한 양적 증가를 표현할 때 '기하급수' 용어를 사용한다. 기하급수라는 것은 체스전설에서와 같이 헤아릴 수 없는 양적 증가를 의미한다.

과학과 연구 데이터

요즘은 기하급수적으로 생산되는 데이터가 화두다. 작은 센서로 부터 거대 실험 장치들이 원유라 불리는 데이터를 실시간으로 생산하고 있다. 소위 4차 산업혁명 시대가 도래했다 말한다. 과학계에서도 데이터가 '뜨거운 감자'다. 중요하지만 쉽지 않은 문제다. 오픈 액세스, 오픈 데이터, 오픈 리포지터리 등으로 구체화되는 오픈 사이언스 운동에서도 데이터가 그 핵심에 있다. 학술 연구 영역에서는 이러한 데이터를 '연구 데이터'라 부른다. 공적 편딩을 받아 수행된 연구 산출물에 대해 법적, 제도적, 기술적 장벽 없이 접근과 활용이 가능하도록 하자는 오픈 사이언스 운동은 주요 선진국에서 다양한 형태의 활동을 이끌어내고 있다. 연구비 지원기관과 도서관, 출판사, 연구자들이 이러한 활동에 참여하고 있다.

과학이란



- (Adopted by Council on November 14, 1999)
- Science extends and enriches our lives, expands our imagination and liberates us from the bonds of ignorance and superstition. The American Physical Society affirms the precepts of modern science that are responsible for its success.
- **Science is the systematic enterprise of gathering knowledge about the universe and organizing and condensing that knowledge into testable laws and theories.**
- The success and credibility of science are anchored in the willingness of scientists to:
 1. **Expose their ideas and results to independent testing and replication by others. This requires the open exchange of data, procedures and materials.**

〈그림 2〉 과학과 오픈 사이언스

본고에서는 이런 활동 중에서 연구 데이터 관리(Research Data Management, 이하 RDM이라 함)와 데이터 관리 계획(Data Management Plan, 이하 DMP 라 함), 데이터 저널과 논문을 둘러싼 다양한 이해 관계그룹의 활동들을 살펴보고자 한다.

한편 과학이란 ‘우리가 사는 세상에 대한 지식을 모으고 조직해서 확보한 지혜를 검증 가능한 규칙과 이론으로 만드는 것’을 의미한다. 결국, 과학이 과학답기 위해서는 연구 과정에서 생산되거나 수집된 데이터와 연구 절차, 연구 자료 등을 공개하고 재사용할 수 있도록 해야 한다. 연구의 재현성(reproducibility)과 복제성(replicability), 일반화 가능성(generalizability)이 확보 되어야 신뢰할 수 있는 연구라 할 수 있다 [2]. 이러한 것이 가능하기 위해서는 연구 과정에서 사용된 연구 데이터 공개와 재사용이 보장되어야 한다. 해외 주요국에서는 국민의 세금이 투입되어 생산된 연구 결과물들(학술 출판물, 연구 데이터 등)을 국가와 기관의 자산으로 인식하고 DMP를 연구자들에게 요구하고 있다. 또한 DMP에 담긴 내용을 실행하기 위해, 연구자 소속 기관의 도서관에게 RDM 서비스를 요구하고 있다.



〈그림 3〉

연구 커뮤니티에서도 이상과 관련된 자발적인 움직임이 오래전부터 있었다. 하지만 연구 데이터 보존과 재사용에 대한 필요성이 커뮤니티에서 논의되고 있었지만 그 실효성은 기대에 미치지 못했다. 네이처 출판사에서 발행되는 ‘Genetics’ 학술지의 경우, 대량의 데이터 세트와 계산도구가 필요한 연구 분야 논문이 실린다. 병렬기술을 포함하는 실험들이 잠재적인 수많은 문제를 안고 있었기 때문에 연구에 사용된 전체 데이터제출이 2001년에 권고 되었다. 하지만 2005년 부터 2006년 까지, Nature Genetics에 실린 18편 논문

중, 두 편의 논문만 연구 재현이 가능하였다. 원시 데이터(raw data) 부재가 주요 원인이었다. 심지어 데이터 재분석 결과 논문의 결론과 다른 결론이 도출되는 사례도 자주 발생하였다 [3]. 신뢰성을 확보하는 연구를 위해서는 원본으로서의 연구 데이터가 반드시 필요함을 의미하는 대목이다.

연구 데이터 동향과 전문 인력

최근 NMC는 향후 5년간(2017년~2021년) 주요 흐름이 될 수 있는 6개의 핵심 트렌드와 6개의 주요 도전, 6개의 기술 발전에 대해 다룬 보고서, 'NMC Horizon Report: 2017 도서관 에디션[4]'을 발표하였다. 보고서에서는 향후 1-2년 내 '연구 데이터 관리'를 위한 새로운 기술 도입이 필요하다고 분석하였다. 또한 향후 3-5년 이내 도서관 공간의 재구성을 위한 기술 도입이 필요하며, 장기적으로 학술 기록의 진화에 대비한 신기술 도입이 필요하다고 분석하였다. 보고서에 따르면 연구데이터는 2014년부터 핵심 트렌드로서 지속적인 관심을 받고 있다.

한편 북미연구도서관협회(Association of Research Libraries, ARL)는 '연구데이터 관리 서비스'를 주제로 보고서[5]를 발행했다. ARL 회원 도서관을 대상으로 각 도서관의 연구데이터 관리와 관련된 활동, 액세스 및 아카이빙 등에 대해 조사한 결과를 담고 있다. 또한 연구데이터 정책, 데이터 보유 정책, 데이터 관리 계획 도구, 작업 기술서, 데이터 보존 웹 페이지, 인력자원 등 각 도서관에서 수행하고 있는 서비스 체제 및 자원 등에 대한 내용을 포함하고 있다. 보고서에서는 다음과 같이 연구 데이터와 관련된 업무를 수행하는 인력들에 대한 설명을 포함하고 있다.

- Data Management Consultant (존 홉킨스 대학)
- Research Data Librarian (메릴랜드 대학)
- Social Sciences Research Liaison Librarian (오타와 대학)
- Digital Library Software Developer (퍼듀 대학)
- Digital Data Repository Specialist (퍼듀 대학)
- Data Curation Librarian (테네시 대학)
- Data Consultant (버지니아 대학)

위에 기술된 다양한 직군들이 처리하는 핵심 콘텐츠는 데이터이다. 과거 전통적인 도서관 서비스 콘텐츠(저널, 논문, 보고서, 단행본, 특허, 동향 등)에 더해, 도서관 서비스 콘텐츠 대상에 데이터 까지 포함되고 있는 것이다. 아직까지 국내 도서관에서는 위와 같은 직군의 인력이 공식적으로 활동하고 있는 사례는 발견하지 못했다. 하지만 해외 주요국의 사례와 같이 연구 데이터와 관련된 소속 기관 연구자들의 자발적 요구에 의해 상기의 직군들이 요구될 것이다.

데이터 저널과 논문

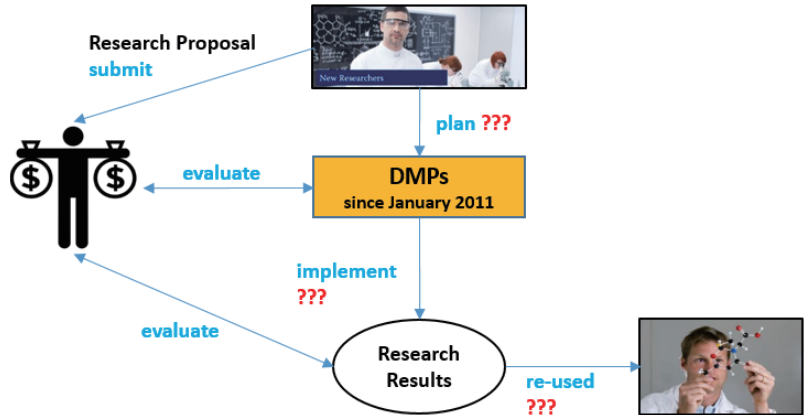
데이터 저널에 투고되는 논문은 독창성이 요구되지 않는다. 또한 중요한 분석 및 결과를 요구하지도 않는다. 데이터 논문은 데이터 검색이 가능하도록 만들며 활용 가치가 높은 고품질 데이터를 보장하며, 데이터를 생산한 연구자 및 공헌자에게 크레딧(credit)을 제공한다. 데이터 논문은 데이터 세트에 대한 설명을 제공한다. 예를 들어 데이터 세트 수집, 처리과정 및 방법, 소프트웨어, 파일 형식 등의 정보를 제공한다.

데이터 출판은 두 가지 방법으로 가능하다. 이 상과 같이 데이터 저널에 데이터 논문을 투고하는 것으로 데이터를 출판할 수 있다. 또 하나의 방법은 데이터 리포지터리에 출판하는 것이다. DCMI (Dublin Core Metadata Initiative)에 정의되어 있는 ‘출판(publish)’은 웹상에서 접근 가능한 상태로 만드는 것을 의미한다. 결국 데이터 저널을 이용하는 데이터 리포지터리를 이용하는 데이터 출판이라는 것이 데이터를 웹에서 접근 가능한 상태로 만드는 것이기 때문에 모두 데이터 출판인 것이다. 최근 해외에서는 Web of Science 서비스를 제공하는 상업회사가 Data Citation Index(이하 DCI라 함) 상품을 개발해 서비스를 제공하고 있다. DCI 데이터베이스에 기관형 리포지터리를 등록해 연구자들이 생산한 데이터와 전통적인 논문서비스를 연계하는 것도 새로운 서비스 패러다임으로 자리매김 하고 있다. 따라서 전문 도서관이나 대학도서관은 현재 운영중인 리포지터리를 데이터 리포지터리로 확장하고 DCI 데이터베이스에 등록되기 위한 노력이 필요하다.



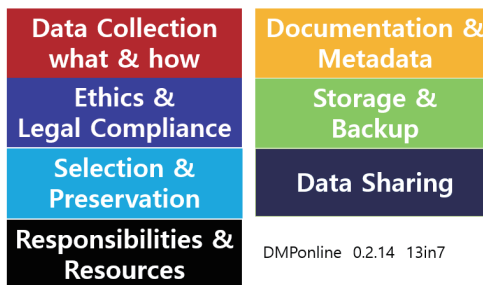
〈그림 4〉 데이터 저널

DMP와 RDM 관계



〈그림 5〉 DMP와 RDM 관계 (도서관 관계 제외)

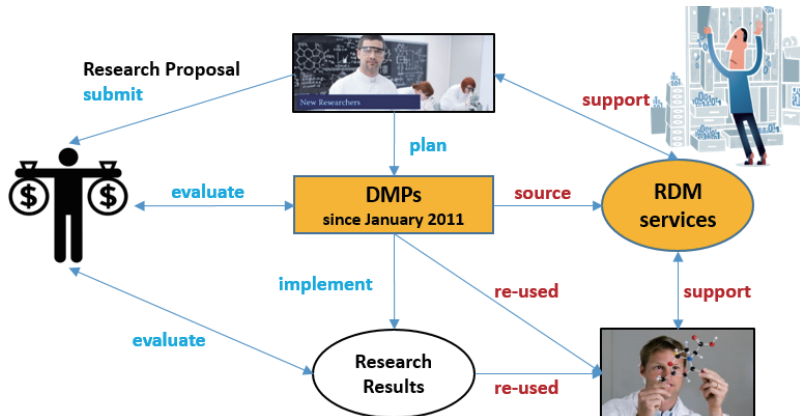
연구 데이터가 국가와 기관의 자산이라는 인식이 확산되면서 체계적인 데이터 관리와 재사용을 보장하기 위해 〈그림 5〉와 같이 DMP가 요구되기 시작했다. 연구자는 연구제안을 위해, 연구비 지원 기관에 연구 계획서와 함께 DMP 제출을 요구 받게 되었다. DMP 항목은 요구하는 기관마다 그 내용에 차이가 있지만, 대부분 〈그림 6〉과 같은 항목들로 구성되어 있다.



〈그림 6〉 DMP 주요 항목

DMP에는 연구 과정에서 어떤 연구 데이터를 수집하고 생산할 것인지? 데이터의 유형(type)과 형식(format)은 무엇이고 크기는 어떻게 되는지? 데이터의 장기적 접근과 사용을 고려해 내용을 작성해야 한다. 또한 데이터 사용과 관련된 윤리적인 문제나 법률적 준수여부도 작성해야한다. 데이터 선정과 보존과 관련된 내용 또한 기술해야 한다. 연구 과정에서 수집, 생산되는 모든 데이터(raw data, processed data 등)를 대상으로 할 것인지? 데이터 선정 기준은 무엇이며, 선

정된 데이터의 보존은 어디에 어떻게 할 것인지?를 기술하게 되어 있다. 또한 연구데이터 재사용성을 보장하기 위하여 문서작업은 어떻게 할 것인지? 데이터를 설명하는 메타데이터 표준은 무엇을 사용할 것인지? 기술하도록 되어있다. 데이터 저장과 백업은 어디에 어떻게 할 것인지? 데이터 공유를 위한 데이터 출판은 어떻게 할 것인지? 이러한 일련의 활동을 위한 자원 확보 및 자원 운영에 대한 책임은 누구에게 있는지? 필요한 자원으로는 무엇이 있는지? 기술하도록 요구받는다. 생각해 보자. 연구자들이 이러한 내용을 모두 기술할 수 있는가? 당연히 연구자는 소속 기관 도서관에 요구를 하게 되었고, 도서관에는 이상의 내용을 지원하기 위한 RDM 서비스를 연구자에게 제공하게 되었다. <그림 7>은 <그림 5>에 도서관의 역할을 추가한 것이다. 그림에서는 연구자(이용자)의 요구가 도서관에 전달되고, 도서관은 RDM 서비스를 연구자에게 제공하는 것을 보여준다. 도서관은 기존에 운영해오던 IR(institutional repository)를 데이터 리포지터리로 확장을 하고 있으며, 데이터 보존과 장기적 접근을 보장하고 있다. 앞에서 DMP 내용을 설명했듯이 연구 데이터 보존과 재사용을 위한 다양한 요구사항이 DMP에 담긴다. 따라서, 도서관에서는 DMP를 분석하여 데이터 보존과 재사용을 보장하기 위한 연구자의 요구사항을 수렴할 수 있다. 이것은 도서관 신규 서비스 개발과 추가적인 개선을 위한 이용자 요구사항 도출이 DMP를 통해 가능하다는 것을 의미한다. Rolando 외(2015)는 연구자들이 작성한 DMP를 분석함으로써, 연구자들이 무엇을 요구하는지? 데이터에 대한 연구자들의 인식 수준은 어느정도인지? 도서관이 새로운 RDM서비스를 어떻게 기획해야 하는지 아이디어를 얻을 수 있으며, DMP가 RDM 서비스 기획의 시작도구로 사용될 수 있다고 주장했다 [6].



<그림 7> DMP와 RDM 관계 (도서관 관계 포함)

연구데이터 설문조사

여기에서는 연구 데이터와 관련된 여러 기관의 설문조사 결과를 살펴본다. 레스터 대학 RDM 설문조사[7] 내용을 살펴보면 연구자들은 RDM 서비스 중 교육과 보존, 저장소 공간, 타 플랫폼과의 연동기능 등을 우선적으로 요구하고 있다. 또한 연구자들은 자신들의 연구 데이터를 설명할 적합한 메타데이터를 가지고 있지 않다.

영국의 6개 고등교육 기관 연구자 약 1천명을 대상으로 한, 설문조사[8] 결과를 통해 다음 사항을 알 수 있다. 디지털 연구 데이터 유형을 조사한 결과 문서 및 PDF 등의 데이터가 75% 이상을 차지하며, 다양한 연구 데이터 생산수단 및 데이터 형태가 존재한다. 또한 연구자들은 데이터 보존 장소로서 IDR(Institutional Data Repository, 이하 IDR이라함)과 연구 데이터 장기 보존을 요구하고 있다. 설문에 응답한 연구자 1,185명 연구자 중에 68% 응답자는 이미 연구 데이터를 공유하고 있거나 미래에 그러하기를 희망한다고 응답하였다. 이러한 응답 요인으로는 '연구 데이터를 공적 자산으로 인식'과 데이터 재사용, 연구 결과 검증, 데이터 인용, 연구비 지원기관 요구, 기관 데이터 정책 순으로 응답하였다. 한편 응답자 중 64% 연구자들은 기관의 연구 데이터 서비스를 사용하고 있지 않았으며, RDM 서비스 존재 자체를 모르는 연구자도 35%에 달했다.

2017년에는 유럽 연구도서관협회(LIBER)에서 학술연구도서관들을 대상으로 연구 데이터 서비스에 관한 설문 조사가 진행되었다 [9]. 22개국 119개 기관이 응답하였다. 설문 결과를 분석한 결과, 도서관은 현재 기술형 서비스(데이터 저장 등) 보다 상담형 서비스(어떻게 정보를 찾는지, 데이터 관리 계획, 메타데이터 표준, 데이터 인용 등)를 제공하고 있다. 연구데이터 서비스 정책을 가지고 있는 기관은 아직 절반 이하에 불과하며, 응답한 도서관장이나 정보센터 책임자 중 2/3가 도서관의 연구데이터 서비스 필요성에 강하게 동의하고 있다고 조사되었다.

한편, 2016년에 캐나다연구비지원기관연합 (Canadian federal funding agencies) 에서 DMPs 개발을 지지하고 데이터가 사용가능하도록 만들기 위해서, 'Tri-Agency Statement of Principles on Digital Data Management' 를 도입하였다. 이에 대응하기 위해서 British Columbia 대학의 사서들은 다음과 같이 3년에 걸친 분야별 설문조사를 수행하였다.

- The Sciences and Engineering (fall 2015)
- The Social Sciences and Humanities (fall 2016)
- The Health Sciences (spring 2017) [10]

2017년 조사에 따르면, 보건학 분야의 연구 데이터 규모는 1GB 미만 47.2%, 50GB 미만 62.9%로 조사되었다. 대부분의 데이터가 문자와 숫자 데이터(70%)로 구성되었다. 응답자들의 75.3%는 데이터 공유가 협력적 학문 연구에 도움이 된다고 응답했으며, 68.5%는 다음세대 연구자들에게 도움이 된다고 응답하였다. 80% 연구자들은 DMP 작성 교육을 위한 워크숍이 필요하다고 했으며, 응답자 85%는 개인 컨설팅이 필요하다 응답했다.

한편, 2017년 유럽 연구도서관협회(LIBER) 설문조사를 통해, LIBER 소속 기관에서는 RDS 기술개발을 위해 도서관 직원들에게 다양한 교육을 제공하고 있음을 알 수 있다. 도서관 직원들에게 컨퍼런스나 워크숍에 참여기회를 제공(78.2%)하고 RDS관련 과정을 이수하도록 지원(59.8%)하고 있다.



〈그림 8〉 유럽 연구도서관협회 소속기관에서 RDS 기술개발을 위한 도서관 직원교육 제공 현황

결론

Resources for Arts, Humanities, Social Sciences, Historical Sciences, Laws & IOE disciplines

Data Management guidelines | Metadata | Data repositories | Ethics

- DataTrain: Open Access Post-Graduate Teaching Materials in Managing Research Data in Archaeology
- ICPSR: Guidelines for
- UK Data Service: Dat
- VADS4R project: Gua

특정 주제 분야 데이터 관리 가이드라인

Data Management guidelines | Metadata | Data repositories | Ethics

- Archaeology Data Service / Digital Antiquity: guidance on project metadata for archaeology
- Data Documentation Initiative (DDI): An international standard for describing statistical and social science data. It is freely available and describes data that result from observational methods in the social, behavioural, economic, and health sciences
- NeSTAR Publisher: A data management repository that can be used for the preparation of data and metadata
- QuEX - Qualitative E by the UK Data Archive for describing qualitat
- UK Data Service - Do..... UK Data Services' catalogue records and an example extract.

특정 주제 분야 메타데이터

Data Management guidelines | Metadata | Data repositories | Ethics

- Cessda (Consortium of European Social Science Data Archives): Brings together social science data archives from across Europe.
- European Union Open Data Portal: A single point of access to a growing range of data from the institutions and other bodies of the European Union
- Europeana Library: A multi-lingual repository of visual, audio, video and written sources from across Europe
- Qualitative Data Re: used in qualitative and mult
- UK Data Service Re: 포지터리 리스트 제공
- UK Data Service Re: funded by the ESRC to support researchers, teachers and policymakers who depend on high-quality social, historical and economic data
- UCL Library Services' Guide to Sources of Socio-Economic Data: Library leaflet detailing primary governmental data websites and UCL's subscriptions to financial & business sources and data archives. Includes tips on using social science data, citation and contacts for further assistance.

Data Management guidelines | Metadata | Data repositories | Ethics

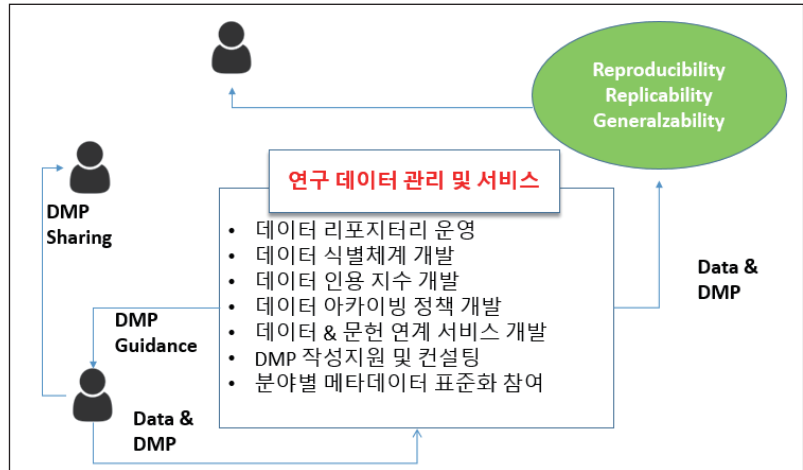
Ethical guidelines and legislation for IOE disciplines

UK Data Service - Legal and Ethic: Includes extensive information on 1 데이터 윤리 정보

access control and further resource.

〈그림 9〉 University College London 대학의 RDM 서비스 사례

지금까지 과학의 특성과 과학 활동에서 연구 데이터 의미를 살펴보았다. RDM과 DMP, 연구 데이터 설문조사 내용을 중심으로 연구 데이터 동향도 살펴보았다. 열린과학(오픈 사이언스) 활동이 글로벌하게 지지받고 있다. 이와 아울러 연구 데이터도 국가와 기관의 자산이라는 인식이 확산되면서 공개접근(오픈 액세스) 운동과 데이터 공개운동이 활발해지고 있다. 이러한 사상과 운동을 제도적으로 뒷받침하는 DMP 작성이 연구자들에게 요구되고 있으며, 이를 지원하기 위해 새로운 도서관 서비스 영역이 요구되고 있다. 바로 RDM 서비스가 요구되는 것이다. 이미 2014년부터 도서관에서 연구 데이터를 차세대 핵심 콘텐츠로 인식하고 있으며, 가까운 시일 안에 이를 처리하기 위한 기술요소들이 개발될 것이라 예측되어지고 있다. 도서관은 기존에 운영하던 IR을 IDR로 확장해 나가고 있으며, RDM 서비스를 위한 다양한 의사결정에 참여하고 있다. 이것은 조직내에서 도서관과 사서의 지위가 높아져가고 있음을 의미한다. 하나의 조직이 연구자들에게 완벽한 RDM 서비스를 제공할 수 없다. 연구자들도 자신들의 연구 커뮤니티에서 오래전부터 사용 중인 인프라를 그대로 사용하고자 할 것이다. 외부시스템과의 연계를 연구자들이 요구한다는 설문조사결과를 보아도 이것을 알 수 있다. 〈그림 9〉는 칼리지 런던 대학의 RDM 서비스 모습을 보여준다. 특정 주제분야의 데이터 관리 가이드라인을 제시하고, 특정 주제분야 메타데이터를 안내하고 있다. 주제분야별 데이터 리포지터리 리스트를 제공하고 있으며, 데이터 윤리 문제도 다루고 있다. 대부분의 서비스 내용이 아웃링크로 구성되어 있다. 자원과 서비스 연계가 RDM 서비스에 사용되는 좋은 예시이다.



〈그림 10〉 연구 데이터 관리 및 서비스 제안

마지막으로 도서관에서 고민해야하는 연구 데이터 관리와 서비스 요소를 살펴 보고 이 글을 마칠까 한다. 〈그림 10〉은 위에서 언급한 설문조사 내용을 중심으로 도서관 RDM에서 필요한 요소를 도식화 한 것이다. 첫 번째 요소로서 IDR 운영이다. 연구의 중심도구로 데이터가 사용되고 있다. 연구자들은 연구에 사용하는 기존 학술정보에 더해, 데이터를 핵심 콘텐츠로 요구하고 있다. 또한 데이터를 확보해 연구 재현과 복제, 일반화 가능성을 높이고자 한다. 이를 지원하기 위해서 데이터 리포지터리 구축 운영이 필요하다. 물론 향후에 DCI 서비스를 활용한 데이터 생산자 및 공헌자에게 혜택을 주기 위해서도 데이터 리포지터리가 필요하다. 이와 아울러 데이터 식별체계 및 인용 지수 개발도 필요하다. 식별 체계는 디지털 객체 식별자(DOI) 적용을 고려해 볼 수 있다. 한국과학기술정보연구원(KISTI)에서 DOI prefix를 부여받아 식별체계에 활용하면 가능하다. 인용 지수의 경우, DCI를 활용한 지수개발을 고려해 볼 수 있다. 두 번째 요소는 데이터 아카이빙 정책개발이다. 도서관의 전통적 서비스 콘텐츠인 도서를 폐기하기 위해서도 정책이 개발되어야 한다. 이와 유사하게 데이터를 아카이빙하기 위한 정책을 개발해서 기관 연구자들에게 홍보를 해야 한다. 이 경우, 구글드라이브나 드랍박스와 같은 상업 서비스와 비교해, 기관의 아카이빙 서비스가 어떤 장점이 있는지를 연구자가 쉽게 이해할 수 있도록 해야 한다. 세 번째 요소는 데이터와 문헌 연계 서비스 개발이다. 연구 데이터를 가장 잘 설명하는 것은 연구 논문과 보고서이다. 이러한 문헌중심의 콘텐츠는 그 유통 생태계가 성숙해 있다. RDM 서비스는 논문을 검색

점점으로 시작해 데이터에 접근하거나, 그 반대로 데이터를 점점으로 시작해 관련 논문에 접근 할 수 있도록 해야 한다. Elsevier 출판사와 PANGAEA 데이터서비스 플랫폼은 이러한 연계서비스 사례를 보여주고 있다. 네 번째 요소는 DMP 작성 지원 및 컨설팅이다. 설문조사에서 드러난 것처럼 연구자들은 DMP 작성에 많은 어려움을 겪고 있다. 따라서 데이터 사서가 DMP 작성을 지원하고 있으며, 데이터 보존, 접근, 출판과 관련하여 실질적 도움을 주는 컨설팅을 제공하고 있다. 마지막으로 요소로는 메타데이터 표준화 참여이다. 하나의 조직이라 하더라도 연구 분야가 다양하기 때문에 연구 데이터를 설명하기 위한 메타데이터가 분야별로 존재할 수 밖에 없다. 데이터 사서는 연구 분야별로 사용되고 있는 기 표준에 대한 정보를 알고 있어야한다. 또한 메타데이터 표준화가 진행되지 못한 분야의 연구자에게는 유사분야의 메타데이터 표준이나 DCMI(Dublin Core Metadata Initiative)에 정의된 메타데이터 요소들을 제안할 수 있어야 한다.

지금까지 연구 데이터와 관련된 국내·외 주요 동향을 살펴보고 도서관 현장에서 무엇을 준비해야 하는지 알아보았다. 연구 데이터와 관련된 거대한 물줄기가 가까운 시일안에 우리나라 도서관 현장에도 밀려올 것이다. 도서관 현장 뿐 아니라 교육 현장에서도 연구 데이터에 관심을 갖고 많은 준비를 해야 할 때다. 끝으로 이 글이 도서관 정책 및 운영계획 수립에 기초 자료로 활용되기를 바란다.



참고문헌

1. 김선태. 2013. 빅데이터 시대 : 도서관과 사서의 새로운 역할, KSLA Bulletin, pp.36 - 44
2. REPRODUCIBILITY, REPLICABILITY, AND GENERALIZATION IN THE SOCIAL, BEHAVIORAL, AND ECONOMIC SCIENCES https://www.nsf.gov/sbe/SBE_Spring_2015_AC_Meeting_Presentations/Bollen_Report_on_Replicability_SubcommitteeMay_2015.pdf
3. Courtesy of C. Glenn Begley, John P.A. Ioannidis. 2015. Reproducibility in Science / Improving the Standard for Basic and Preclinical Research. 2015 Circulation Research / <https://doi.org/10.1161/CIRCRESAHA.114.303819>Circulation Research. 2015;116:116-126 Originally published December 31, 2014
4. Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Giesinger Hall, C., Ananthanarayanan, V., Langley, K., and Wolfson, N. (2017). NMC Horizon Report: 2017 Library Edition. Austin, Texas: The New Media Consortium.
5. SPEC Kit 334 <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>
6. Rolando, L. , Carlson, J. , Hswe, P. , Parham, S. W., Westra, B. and Whitmire, A. L. (2015), Data Management Plans as a Research Tool. Bul. Am. Soc. Info. Sci. Tech., 41: 43-45. doi:10.1002/bult.2015.1720410510
7. RDM survey: overview findings Ian Rowlands University Library <https://www2.le.ac.uk/services/research-data/advice-and-support/survey>.

8. Johnson, Rob; Chiarelli, Andrea; Parsons, Tom (2016): Data asset framework (DAF) survey results 2016. figshare. Fileset. <https://doi.org/10.6084/m9.figshare.3796305.v4>
9. Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. and Allard, S., 2017. Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), pp.23 – 44. DOI:<http://doi.org/10.18352/lq.10180>
10. Three UBC Research Data Management (RDM) Surveys: Science and Engineering, Humanities and Social Sciences, and Health Sciences. Summary Report. July 2017.