



연구데이터 관리 현황 및 향후 계획: 한국철도기술연구원을 중심으로

남은경 _ 한국철도기술연구원

1. 서론

연구자들은 연구를 수행하면서 연구 목표에 따른 최종 결과를 도출하기 위하여 각종 연구 활동을 수행하고, 각 연구 단계마다 다양한 형태의 연구데이터, 정보, 지식을 생성한다. 이러한 데이터, 정보, 지식은 연구의 질 향상을 위해서 중요한 자원이나, 수집에 어려움이 있다. 이에 도서관들은 효율적인 수집 및 공유를 위하여 기관 내에서 다양한 노력들을 하고 있다.

본 논문에서는 한국철도기술연구원(이하 연구원)에서 생산하는 연구데이터의 현황, 연구데이터 수집 및 서비스 현황, 향후계획에 대해서 살펴보고자 한다.

2. 연구데이터 정의

연구데이터(research data)는 학자 및 기관에 따라 다양하게 정의되고 있고, 과학데이터와 혼용하여 사용되고 있다.

김선태(2010)는 과학데이터란 “연구자의 연구 활동 과정 중 생성되는 다양한 유형의 사실적 기록을 의미한다. 즉, 연구 활동을 통하여 생산된 연구 활동의 기록물로서 관측, 감시, 조사, 실험, 분석, 계산 등의 과정을 통하여 생산된 문자, 이미지, 오디오, 동영상 등의 아날로그 및 디지털 형식을 포괄하는 데이터”라고 정의하고 있다[1].

OECD(Organization for Economic Co-operation and Development)에서는 연구데이터란 “수치, 문자, 이미지나 음성 등의 사실적 데이터이며, 주로 과학 연구의 주요한 원천으로 사용하고, 과학분야 학술자 사이에서 연구결과를 검증하기 위해 필수적인 것”이라 규정하고 있고 연구데이터에 포함되지 않는 연구

자료를 별도로 명시하고 있다. 여기에는 연구노트(laboratory notebooks), 예비 분석(preliminary analyses), 논문의 초안(drafts of scientific papers), 연구계획서(plans for future research), 동료리뷰(peer reviews), 동료들 간의 개인적인 커뮤니케이션(personal communications with colleagues), 물리적인 대상(physical objects) 등이 포함된다[2].

김은정(2012)은 연구데이터가 “연구수행과정 중에 산출되는 데이터로서 원 데이터와 분석된 2차 자료를 포함하며, 관찰, 조사, 실험, 경험에 기반하는 사실로 수치, 텍스트, 오디오, 컴퓨터 파일 등 다양한 형태의 데이터를 포함한다”라고 정의하였다[3].

이와 같이 연구데이터에 대한 정의는 학자들마다 다르고, 본 논문에서의 연구데이터는 연구의 최종결과물인 연구보고서, 2차 자료, 시험데이터 등을 포괄하는 의미이다.

3. 연구데이터 관리 현황

1) 연구데이터 생산 현황

「과학기술분야 정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 의거 설립된 한국철도기술연구원은 프로젝트성 연구사업을 수행하는 기관이다. 연구원에서는 각종 보고서, 직원 발표 및 게재 논문, 시험데이터 등이 생산되고 있으며 세부 내용은 <표 1>과 같다

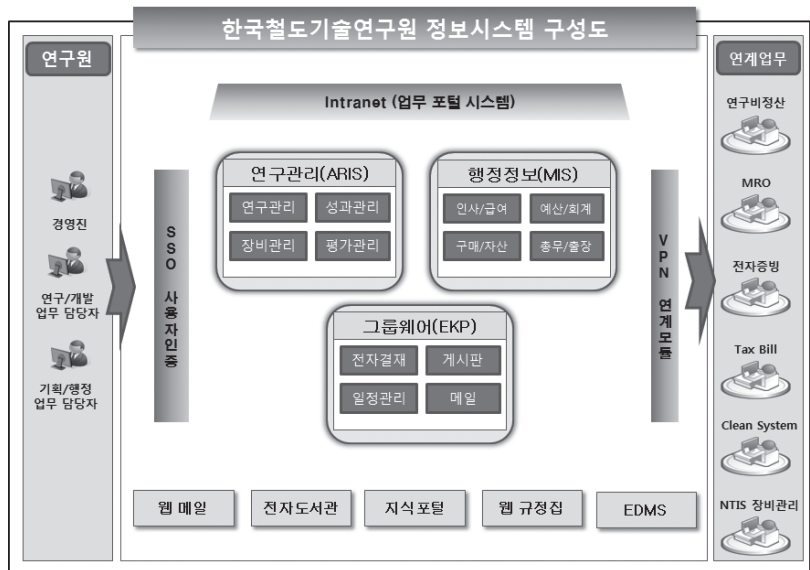
<표 1> 연구원 생산 연구데이터의 종류

분류	내용
연구보고서	연구사업의 수행결과로 발생한 보고서
위탁연구보고서	연구사업의 일부과업을 외부 기관에 위탁하고 위탁연구 결과를 수록한 보고서
용역보고서	구매계약을 통해 이루어지는 용역건으로 용역결과를 수록한 보고서
장기자문보고서	외부전문가 장기자문 종료 후 자문 결과를 수록한 보고서

연구논문	연구개발의 성과 또는 결과를 학술지, 학술대회 등에 일정한 형식을 갖추어 발표하는 글
국외출장보고서	국외출장 결과를 일정한 서식에 따라 작성, 보고하기 위한 보고서
시험 및 해석 데이터	연구수행 중 실시되는 시험 또는 해석 데이터

2) 연구데이터 수집 현황

연구원의 정보시스템은 <그림 1>과 같이 행정정보시스템, 연구관리시스템, 전자도서관시스템으로 크게 나누어진다. 연구원에서 생산되는 각종 연구데이터는 <표 2>와 같이 업무성격에 따라 별도의 시스템에서 관리되고, 각각의 시스템은 유기적으로 연계되어 있다. 도서관은 연구를 수행하면서 생산되는 각종 연구데이터가 각각의 업무프로세스에서 자동으로 시스템에 등록되도록 체계를 구축하였고, 각 시스템에 등록된 연구데이터는 시스템 연계를 통하여 전자도서관시스템에 통합적으로 축적된다.



<그림 1> 연구원 전체 시스템 구성도

〈표 2〉 시스템별 연구데이터 현황

분 류	내 용
연구관리시스템	· 대상: 위탁연구보고서, 장기자문보고서 · 연구자가 위탁연구보고서 및 장기자문보고서 연구관리시스템 등록, 전자도서관은 시스템 연계를 통하여 메타정보 및 파일 전송
행정정보시스템	· 대상: 용역보고서, 국외출장보고서 · 연구자 검수시 용역보고서 등록, 국외출장 복명시 국외출장보고서 등록, 전자도서관 시스템 연계를 통하여 메타정보 및 파일 전송
전자도서관시스템	· 대상: 연구보고서, 연구논문 · 연구자 연구보고서 및 연구논문 전자도서관시스템을 통하여 등록 연구논문 연1회 평가시스템에 메타 전송
기타	· 시험 및 해석데이터: 팀 또는 개인 단위로 관리

(1) 연구보고서

연구보고서는 연구수행의 결과를 수록하고 있는 자료이다. 연구자들은 사업이 종료되면 도서관에 연구보고서를 지정된 기간 이내에 제출하여야 한다. 그러나 연구사업 평가가 지연되거나 발주처의 요구에 의거 발간이 지연되는 경우가 종종 발생되고, 도서관에서는 연구보고서 발간이 지연되는 사유의 확인이나 연구보고서 수집에 어려움이 있었다.

이에 2017년 연구관리규정 개정을 통하여 연구보고서를 도서관에 미제출시 연구책임자의 일부권한을 제한할 수 있는 근거를 마련하였고, 2018년 본격적으로 시스템 개편을 통하여 제도를 시행하였다.

개선된 제도에서 연구자들은 연구관리규정에서 지정한 기간 이내에 도서관에 연구보고서를 제출하여야 하고, 미제출시 그룹웨어 메인화면에서 미처리 알림 제공 및 연구책임자의 일부 행정행위를 제한한다. 미제출 사유가 발생한 연구사업의 경우는 부서장 승인 후 제출기간을 연장할 수 있다.

도서관에 수집된 연구보고서는 연구원 내부 위원회에서 결정한 공개구분에 따라 연구원 내외부에 서비스된다.

(2) 위탁연구보고서

연구원에서는 연구원이 주관으로 수행하는 연구사업의 일부 과업을 외부기관에 위탁하고, 위탁연구보고서는 위탁연구사업의 결과를 수록한 보고서이다. 위탁연구보고서는 사업이 종료되면 도서관에 제출하도록 규정화되어 있다. 도서관은 연구사업 종료 후 연구책임자 완료보고서 보고서 첨부을 의무화하여 시스템에 축적될 수 있는 구조를 만들었다. 연구관리시스템에 등록된 보고서는 시스템 연계를 통하여 전자도서관으로 전송된다.

(3) 용역보고서

용역보고서는 구매계약을 통하여 이루어지는 용역의 결과를 수록한 보고서이다. 연구원에서는 용역보고서 수집을 위하여 2016년부터 시스템에 필수적으로 보고서가 등록되도록 프로세스와 시스템을 개편하였다. 연구자는 용역 발주를 위한 구매요구서 보고서 유무를 체크하고, 체크된 구매요구건은 검수 시 보고서 첨부이 의무로 되어 있다. 전자도서관은 시스템 연계를 통하여 보고서과 메타데이터를 전송 받을 예정이다.

〈표 3〉 용역보고서 제출 기준

분 류	내 용
시험용역	· 시험환경, 시험방법, 시험결과 보고서 수록 · 최종 Raw Data 별도 파일로 등록 (권고)
특허분석	· 사용 DB, 검색식, 분석결과 보고서 수록 · Raw Data 별도 파일로 등록 (권고)
학술용역 동향,실태조사	· 연구보고서 수준으로 작성
시스템 구축	· 매뉴얼 수준으로 작성
시제품제작	· 해석이나 성능평가 포함한 제작일 경우 보고서 제출

(4) 장기자문보고서

장기자문보고서는 외부전문가의 장기자문 종료 후 자문 결과를 수록한 보고서로, 제도적으로는 2016년 12월 모든 장기자문보고서에 대해서 보고서 발간을 의무화하였다. 현재는 장기자문보고서를 관리할 수 있는 시스템을 개발 중에 있고,

2018년 초 오픈할 예정이다. 전자도서관은 시스템 연계를 통하여 보고서와 메타데이터를 전송 받을 예정이다.

(5) 연구논문

연구논문은 내부 직원들이 학회에 게재 또는 발표하는 논문으로, 연구자들은 각종 평가를 위하여 수시로 전자도서관시스템에 연구논문을 입력한다. 도서관은 연구자들이 입력한 정보를 확인한 후 년1회 시스템 연계를 통하여 평가시스템으로 메타데이터를 전송한다.

(6) 국외출장보고서

국외출장보고서는 국외출장 결과를 수록한 보고서이다. 제도적으로 출장자는 귀국 후 국외출장복명을 하도록 되어 있고, 국외출장복명시 국외출장보고서 첨부이 의무화되어 있다. 도서관은 시스템 연계를 통하여 보고서와 메타데이터를 전송받는다.

(7) 시험 및 해석 데이터

연구결과의 원자료인 시험 및 해석 데이터는 증거로서의 가치와 정보로서의 가치를 가지고 있는 중요한 데이터이다. 대형장비들의 경우 대부분 장비시스템에서 시험 데이터가 관리되고 있고, 그 외의 장비나 해석데이터의 경우는 개인 단위 또는 연구사업 단위로 생산·관리되고 있다. 재이용은 연구자들 간에 비공식적인 방법으로 이루어진다.

3) 연구데이터 서비스 현황

연구를 수행하면서 발생하는 각종 보고서 및 논문 등의 연구데이터는 다양한 업무시스템에서 등록되고 시스템 연계를 통하여 전자도서관시스템에서 통합적으로 관리되며 서비스된다.

전자도서관시스템에서는 도서관에서 관리하는 소장도서, 연구데이터, 라이선스를 보유한 전자정보 등 모든 콘텐츠들의 통합검색을 지원하고, 각 유형별 검색을 지원한다. 연구데이터의 경우는 과제별로 연구데이터를 통합해서 관리할 수 있는 검색기능을 제공한다. 또한 모든 연구데이터들에는 기술트리(주제분류)가

부여되어 있고, 별도의 키워드 검색 없이 기술트리별로 브라우징하여 관련 연구 데이터 목록을 확인 할 수 있다. 키워드 검색 후 제한 기능으로 기술트리를 사용할 수도 있다.

연구보고서의 경우는 해당 연구보고서와 관련된 위탁연구보고서, 전년도 또는 차년도 보고서에 대한 연결 서비스를 제공한다. 서비스되는 모든 연구데이터는 공개구분에 따라 내외부에 서비스된다.

4. 향후계획

도서관은 연구자들이 연구를 수행하면서 생산하는 각종 연구데이터들을 업무 프로세스 상에서 등록될 수 있도록 제도와 프로세스를 구축하였다. 그러나 시험 또는 해석 데이터의 경우는 공공재라는 인식의 부족과 제도의 부재로 통합적인 관리가 되고 있지 않다.

시험데이터는 연구성과에 대한 증거로서의 가치와 정보로서의 가치를 동시에 가지고 있는 중요한 정보이다.

이에 도서관에서는 시험데이터를 체계적으로 관리하기 위하여 기술분야별 연구수행 패턴 분석, 메타데이터 표준 및 철도에 적합한 메타데이터(안) 마련, 시험 데이터 수집을 위한 제도 마련 등의 기초 연구를 수행할 계획이다.

참고문헌

1. 김선태, 한선화, 이태영, 김용. 2010. “과학데이터 보존 및 활용모델에 관한 연구” 한 국비블리아학회지, 21(4): p.84.
2. OECD. 2007. OECD principles and guidelines for access to research data from public funding: pp.13-14.
3. 김은정(2012), “연구데이터 수집에 영향을 미치는 요인분석”, 중앙대학교 기록관리 학과 박사학위논문
4. 한국철도기술연구원. 2016. “철도기술 지식경영 촉진방안 및 시행방안 연구; pp24-25,
5. 남은경,김경희,강석주. 2017. “철도분야 연구데이터관리 현황 및 전략, ASIA Data Week,