

# 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구\*

- 기초학문자료센터를 중심으로 -

## Strategies for Improving the Collection and Use of Research Data in the Humanities

심원식(Wonsik Shim)\*\*, 안혜연(Hye-yeon Ahn)\*\*\*, 변제연(Jeayneon Byun)\*\*\*\*

### 목 차

- |   |                              |
|---|------------------------------|
| 1. 서 론                                  | 4. 기초학문자료센터를 통한 인문학 분야 DB 구축 |
| 2. 선행 연구                                | 5. 연구데이터 수집 및 활용성 증진을 위한 제안  |
| 3. 국내 인문사회 연구데이터 관리의 대표 사례:<br>기초학문자료센터 | 6. 결 론                       |

### 초 록

정보기술 및 데이터 네트워크의 발전은 다양한 학술연구에서 생산되는 데이터의 양의 폭발적인 증가 및 활용 가능성을 가져왔다. 하지만 관련 정책 및 인프라 구축의 부재로 인해 연구데이터의 관리 및 공유가 체계적으로 이루어지지 않고 있다. 특히 과학 및 공학 분야와 비교할 때 인문사회 분야에서는 데이터의 체계적인 수집 및 공유를 뒷받침하는 데이터 리포지토리가 취약하다. 본 연구에서는 국내 인문사회 분야 연구데이터 및 자료의 대표적인 리포지토리인 한국연구재단의 기초학문자료센터를 중심으로 인문학 분야 연구데이터의 체계적인 수집, 활용 및 보존을 위한 프로세스 및 전략을 제시하고자 한다. 특히 데이터센터 운영자의 관점에서 효과적인 데이터 수집을 위한 과제분석과 인문학 분야의 특성을 반영한 원자료 정보 구축에 초점을 맞추고 있다. 또한 향후 국내에서 정부의 재정지원을 받아 수행된 연구에서 생산, 산출되는 연구데이터의 수집 및 활용성 증진을 위한 방안으로 데이터관리계획 관련 정책의 시행, 데이터 관리, 수집 및 공유를 위한 교육 및 지원 서비스 제공을 통한 환경 구축, 데이터 전문가의 양성, 연구데이터의 인용을 통한 활용체계 구축 등을 제시하고 있다.

### ABSTRACT

The rapid growth of information technologies and data networks has increased the volume of data generated from scholarly research and the possibilities of re-using and sharing such data. However, there is a serious problem of management and sharing of research data due to the lack of facilitating policies and supporting infrastructure. In particular, few data repositories exist that support systematic collection and sharing of research data in the humanities. In this regard, the Korea Research Memory (KRM) established by the Korea Research Foundation is a rare exception. The purpose of this research is to present specific processes and strategies that can facilitate the data collection, reuse and preservation through the KRM using task analysis and source document gathering as main focal points. In addition, in order for the effective collection and sharing of research data, the following recommendations are proposed: 1) the need for the adoption of data management plan related policies that govern the collection and sharing of research data generated from publicly funded research projects, 2) the need for training and support services for individual researchers and research institutes, 3) the need for training data specialists, and 4) the citation scheme and structure designed for research data.

키워드: 연구데이터, 공유, 한국연구재단, 기초학문자료센터, 인문학

Research Data, Sharing, Korea Research Foundation, Korea Research Memory, Humanities

\* 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014S1A1A1021266).

\*\* 성균관대학교 문헌정보학과 교수(wonsikshim@skku.edu) (제1저자)

\*\*\* 성균관대학교 문헌정보학과 박사과정(gpdus23@gmail.com) (교신저자)

\*\*\*\* 성균관대학교 문헌정보학과 박사과정(bjy0228@gmail.com)

논문접수일자: 2015년 7월 29일 최초심사일자: 2015년 7월 30일 게재확정일자: 2015년 8월 7일  
한국문헌정보학회지, 49(3): 155-183, 2015. [http://dx.doi.org/10.4275/KSLIS.2015.49.3.155]

## 1. 서 론

정보기술 및 데이터 네트워크의 발전은 데이터의 생산, 저장, 공유 및 보존에 큰 변화를 가져왔다. 더 많은 데이터가 더 빠른 속도로 생산될 수 있으며, 이와 동시에 대용량의 데이터를 저장할 수 있는 다양한 매체와 기술 인프라도 발전하고 있다. 하지만 정보기술의 발전이 데이터의 체계적인 관리를 보장하지는 않는다. 더 많은 정보를 손쉽게 생산, 복사하고, 저장할 수 있기 때문에 오히려 일관성 있는 관리가 어려운 측면도 있다. 결과적으로 많은 데이터가 연구자의 개인 컴퓨터나 관리되지 않는 컴퓨터 서버에 방치되는 현상이 계속적으로 누적되어 일어나고 있다.

데이터의 효과적인 수집 및 관리 저해의 주요 인으로 논의되고 있는 것들은 크게 세 가지 정도이다. 첫째, 연구에 대한 보상체계가 모두 학술논문, 학술대회 발표, 저서, 특히 등과 같은 연구 결과물의 생산에만 집중적으로 연계되어 있다. 둘째, 연구결과물의 경우 비교적 정형화되어 있어 이것들을 한 곳에 모으고 조직화하는 것이 상대적으로 용이한 반면 데이터는 내용과 형태면에서 매우 다양하여, 표준형식으로 정형화하기 어렵기 때문에 데이터를 모아서 일관성 있는 방법으로 서비스 하는 것 또한 매우 어렵다. 셋째, 연구데이터의 경우는 단순히 데이터의 저장 및 보존을 위한 기술적인 지원체계뿐 아니라 학술커뮤니케이션의 관행 및 과정에서 데이터의 출판, 등록, 인증, 접근 및 유통 그리고 이용이 되어야 하는데 아직까지 데이터와 관련된 학술커뮤니케이션 관행과 절차가 확고하게 자리를 잡지 못한 상태이다.

국가적 차원에서 연구비 지원기관 및 정책기관에서의 연구데이터 수집 및 공유가 연구성과의 축적과 재이용의 측면에서 중요시되고 있다. 정부연구비의 지원을 받은 과제로부터 생산된 연구데이터는 공공재이고 누구나 접근할 수 있도록 공유해야 한다는 주장이 2007년 OECD 연구데이터 관리와 공유의 기본원칙 중 하나로 제시되었다. 이 원칙에 기초하여 미국, 영국 등의 정부연구비지원기관에서는 지원 과제로부터 산출된 데이터를 데이터 리포지토리를 통해 보존하고 공유 및 공개할 것을 권고하고 있다(김지현 2013). 연구데이터 리포지토리는 크게 연구기관(대부분의 경우에 대학) 또는 연구소에서 운영하는 기관 리포지토리(Institutional Repository)와 특정 학문 분야의 연구데이터를 저장하고 공유하기 위한 분야별 리포지토리(Disciplinary Repository)의 두 가지로 운영되고 있다.

연구데이터의 체계적인 수집과 공유의 수준은 학문 분야에 따라 다른 양상을 보인다. 데이터의 공유가 가장 활발하게 나타나는 분야는 해양 환경, 천체물리학, 생명공학 등이 대표적이다. 이런 분야의 특징은 데이터를 반복해서 수집하기 어렵고, 대규모 장비 및 설비투자를 통해서 수행되거나 데이터 공유의 효과가 크고 직접적이라는 것이다. 사회과학 분야에서도 인구통계 데이터, 체계적인 표본을 대상으로 한 다양한 설문 데이터의 축적이 매우 중요하다. 대표적으로 미국의 The Inter-University Consortium for Political and Social Research(이하 ICPSR), The Harvard-MIT Data Center, 영국의 UK Data Service 등이 수십년 동안에 걸쳐 관련 데이터를 수집, 정리, 제공하고 있다. 국내에서는 한국사회과학데이터센터(KSDC), 한국사회조

사데이터아카이브(KOSSDA) 등이 사회과학 분야 데이터를 아카이빙하고 있다. 인문학 분야에서의 데이터센터 혹은 아카이브 사례로는 인류학 분야의 자료를 구축하고 서비스하는 미국 예일대학의 Human Relations Area File, 캐나다의 민속박물관과 대학 연구소들을 연계하여 캐나다의 다양한 문화민속 자료를 수집하고 연구하기 위해 운영되고 있는 Reciprocal Research Network 등을 사례로 들 수 있고 국내에서는 한국역사정보통합시스템이 역사 분야 자료, 20 세기민중생활사연구소가 관련 분야 자료를 제공하고 있다.

인문학 및 사회과학 분야의 연구자들이 연구 데이터를 생산, 보존하는 시스템에 의존하고 있는 경향이 높아지고 있지만(American Council of Learned Societies 2006), 자연과학분야에 비해서는 학문분야별 데이터의 수집, 관리 및 공유를 위한 제도적, 시스템적 기반이 미약한 실정이다. 인문학 분야는 사회과학 분야에 비해 다양한 학문분야로 구성되어 있으며, 연구 방법론 또한 고전을 비롯한 텍스트 자료의 분석, 실험, 설문 기반 연구, 현장 관찰, 창작 등으로 매우 다양하다. 따라서 인문학 연구에서 사용되거나 생산되는 데이터의 종류도 텍스트 자료뿐 아니라 녹음/동영상, 이미지, 날장자료, 기사자료 등으로 매우 다양하게 산출되고 있다. 이런 상황에서 국내에서는 한국연구재단이 지원한 인문사회분야의 연구과제에서 산출된 다양한 연구구성파일 및 데이터를 수집하고 데이터베이스로 구축하여 연구자를 포함한 정보이용자에게 제공해주는 것을 목적으로 기초학문자료센터(Korea Research Memory, KRM)를 2006년부터 본격으로 운영하고 있다는 점은 주목할

만하다. 지금까지 기초학문자료센터를 중심으로 인문사회분야의 데이터 아카이브에 대한 다양한 논의가 진행되어 왔다(이수상 외 2007; 임건면 2007; 장진규 2007; 고영만 2010; 배경재 2011; 신영란, 정연경 2012). 하지만 본 논문은 기초학문자료센터 사업 초기부터 인문학 분야를 담당하여 자료를 수집, 정리하여 데이터베이스를 구축하는 업무를 수행해 온 저자들이 내부적인 관점에서 인문학 분야의 특수성을 고려하여 실제 연구데이터의 수집, 활용을 위해 사용한 방법론을 소개하고 이를 통한 향후 개선책을 모색해보고자 하였다는 점에서 차별성이 있다. 따라서 본 논문에서는 기초학문자료센터에 대한 개관과 더불어 인문학 분야의 특성을 고려한 업무 프로세스를 과제분석, 원자료 정보 구축의 관점에서 설명한다. 또한 다년간에 걸친 실제 구축 경험을 바탕으로 향후 인문학 분야의 원자료의 원활한 수집 및 활용성 제고를 위한 정책, 인프라 및 활용체계의 구축에 대한 대안을 제시하고자 하였다.

## 2. 선행 연구

연구데이터의 활용성 증진에 영향을 미치는 주요한 요인으로 연구데이터의 수집, 관리 및 공유에 대한 연구자의 인식과 참여, 관련 정책 그리고 운용 시스템(리포지토리) 등을 들 수 있을 것이다. 또한 특정 학문 분야의 연구자에게서 나타나는 특정 정보행위 역시 이러한 요인들과 밀접한 관계를 갖고 있기 때문에 연구데이터의 관리 및 공유 체계에도 그 주요 특성을 반영한 실질적인 적용이 필요하다. 따라서 본 장에

서는 인문학 분야 연구자의 정보행위, 그리고 연구데이터의 관리 및 공유 체계에 관련한 선 행연구들을 살펴보았다.

### 2.1 인문학 분야 연구자의 정보행위

1970년대부터 1990년대 초반까지 약 20년 간 인문학자들의 정보행위를 다룬 문헌의 리뷰 (Stone 1982; Watson-Boone 1994)를 통해 확인된 특징은 인문학자들은 주로 독자적인 작업 (working alone)을 수행하며, 활용자료에 대한 개인적 해석이 연구결과를 도출하는데 있어 중심적인 역할을 한다는 점이다. 사용되는 자료들은 예술작품, 텍스트, 원고, 녹음자료 등으로 다양했으나 주로 단행본과 학술지를 활용하며 이 중에서도 단행본을 더 많이 사용하는 것으로 나타났다. 이러한 자료들은 보다 더 다루기 쉽고 검색이 용이한 환경기반이 갖추어진 전자자원으로 이용빈도가 증가하는 경향을 보였다. 또한 자료의 브라우징을 통해 우연히 발견하게 되는 정보나 아이디어를 매우 중요하게 여겼으며 이러한 방법으로 정보를 모으고 선택·해석하는 과정을 통해 지식으로의 변환이 이루어졌다. 이는 인문학자들이 연구 관련 핵심자료를 갖고 있더라도 그들의 실제 연구경로가 변형적이고 예측불가능하며 다양한 유형의 리딩(reading)을 통해 이것을 집필 단계로 이어가는 것을 확인한 Palmer and Neumann(2002)의 연구를 통해서도 알 수 있다.

이후 2000년대 후반 국내외 인문학자의 정보자원 이용행태에서 나타나는 현상 및 변화를 살펴본 연구들(윤정옥 2009a; 윤정옥 2009b, 11: Tahir, Mahmood and Shafique 2010)에서도

인문학자들은 독자적인 연구행태를 보였고 단행본, 인쇄학술지와 같은 전통적 자원의 사용을 고수함과 동시에 전자원문, 전자학술지 등 텍스트 형태의 전자자원에 대한 의존도 또한 높게 나타났다. 또한 인문학자들은 전자자원의 사용이 연구수행에 도움을 주는 것으로 인식했으며 전자정보원의 접근성 및 이용편의성, 그리고 정보품질 등을 중요하게 여겼다.

이렇듯 정보기술 및 데이터 네트워크의 발전은 보다 풍부한 자원의 생산과 이용을 가능하게 했고, 한 연구자의 연구성과가 보다 빠르게 다른 연구자의 핵심자료로 활용될 수 있도록 하면서 연구의 효율성을 높이는데 주요한 영향을 미쳤다. 이러한 환경에서 연구자들의 연구데이터 보존 및 공유를 위한 노력이 요구되게 되는데, 인문학 분야에서의 데이터 공유는 연구자들에게 친근하지 않은 개념으로, 특히 독자적이고 개인적인 성격을 띠는 연구에서는 데이터 공유가 이루어지기 어려운 부분이 있다. 또한 인문학 분야에서 주요자원으로써 활용되는 자료는 대부분 동일 분야 선행연구자의 데이터로 ‘데이터와 데이터의 근원(data and sources of data)’ 간 경계를 모호하게 하여, 과학 분야나 사회과학 분야에서보다 각각의 연구데이터를 규정, 분류하여 관리하는 것이 쉽지 않다(Jacoby 2010; Burrows 2011).

Schäfer et al.(2011)는 ODE project(Opportunities for Data Exchange)에 참가한 기관의 이해관계자들을 대상으로 인터뷰를 실시하여, 데이터를 공유하고 재이용하는데 있어 추진 요인(동인, 動因)과 장애 요인에 대한 그들의 경험과 관점을 묘사하도록 했다. 이 중 영국의 대표적 인문사회 분야 리포지토리인 UK Data

Archive의 연구 데이터 관리 지원 담당자들은 인터뷰에서 인간과 관련된 데이터는 다른 데이터보다 특별한 처리(handling)를 필요로 하고 해당 데이터에 포함된 개인정보를 익명화하는 것, 데이터 공유나 출판 관련 동의를 확실히 하는 것 등 요구되는 사항이 많으며 이와 관련하여 연구가 진행되는 동안 연구자의 참여가 필수적임을 강조하였다. 그러나 실제 연구데이터 공유에 대해 연구자들은 그들의 데이터를 공유하는 것이 처음인 경우가 많으며, 어떻게 공유해야 하는지를 잘 모르고 단지 데이터 공유 이전에 고려해야 할 몇 가지 중요한 사항들이 있다는 것을 인지하고 있을 뿐, 이에 대한 자세한 내용을 알지 못하고 있는 경우가 많다고 언급했다. 이러한 이유에서 연구자가 자문을 구할 수 있는 기관이나 전문가가 필요하며, 실제 UK Data Archive에서 데이터 공유 측면을 위한 컨퍼런스를 개최한 후 더 많은 이차 분석 연구(secondary analysis projects)가 실시되었고 이를 데이터의 공유와 재이용이 보다 활발해진 것으로 해석했다.

국내 연구자들을 대상으로 하여 연구데이터 관리에 대해 조사한 연구(김지현 2012)에서는 인문사회 분야와 과학 분야(자연과학, 의학 및 공학)의 두 집단 간 연구자가 생산 또는 수집하는 연구데이터 유형에 많은 차이가 있었으나, 두 집단 모두 타인의 연구데이터를 활용한다는 응답자의 비율(약 30%)이 본인의 데이터를 공유한다는 응답자의 비율(약 20%)보다 높게 나타났다. 상당수의 연구자가 타인의 데이터 수집에 연구데이터 리포지토리를 활용하고 있었는데, 특히 데이터의 장기보존 및 메타데이터 작성에 대한 만족도가 낮게 나타나 이를 지원

할 수 있는 연구데이터 아카이빙 서비스의 개발이 필요한 것으로 확인되었다.

최근 수행된 유소영(2015)의 연구에서도 국내 인문학자들이 연구데이터 아카이브를 통해 자료를 획득함으로써 연구 시간 및 비용의 측면에서 효율성 제고가 가능하다고 인식하는 것으로 나타남에 따라, 연구데이터 리포지토리를 인문학 데이터 획득에 적극적으로 활용할 수 있도록 하고 그에 따른 데이터의 공유가 이루어질 수 있는 기반이 마련되어야 함을 저자는 강조하였다.

이처럼 인문학 분야 연구자의 정보행위에 있어 과거부터 오늘날에 이르기까지 일관되게 나타나는 특징은 독자적인 연구 행태로, 단행본과 학술지를 주요 자료로 활용하며, 브라우징을 통해 발견하는 정보나 아이디어가 연구 수행에 중요한 역할을 한다는 점이다. 하지만 이러한 인문학 환경은 연구데이터 관리와 공유를 타 분야에 비해 어려운 상황으로 이끈다. 따라서 많은 연구자들이 연구데이터의 관리 및 공유, 재이용을 위한 연구데이터 리포지토리가 필요하며 이와 관련한 보다 실제적인 운영이 이루어져야 함을 강조하고 있다.

## 2.2 연구데이터 관리 및 공유

연구데이터의 수집과 공유에 있어서는 연구자의 태도 및 참여 정도, 동기부여요인 및 장애요인 등을 살펴본 다양한 연구들이 진행되고 있다.

Hedstrom and Niu(2008)는 연구데이터 제출 의무가 부여된 연구자들을 대상으로 하여 그들의 내적, 외적 동기부여요인이 데이터 제출

관련 태도 및 행위에 어떠한 영향을 미치는지를 설문을 통해 조사하였으며, Perry(2008)는 캐나다 SSHRC(Social Science and Humanities Research Council)의 연구비 지원을 받은 연구자들이 연구 수행 과정에서 생산된 데이터를 얼마나 저장했는지, 그리고 연구데이터의 제출 및 공유, 보존과 관련하여 어떠한 태도와 관심을 나타내는지에 대해 확인하였다.

국내에서는 연구데이터의 체계적인 수집 및 공유의 필요성을 강조하면서 연구데이터 공유의 장애요인으로 데이터 공개에 대해 연구자에게 부여되는 구체적인 인센티브가 없는 점, 관련 정책 및 체제의 부재 등을 언급한 연구(이상환, 심원식 2009), 외국의 연구데이터 관련 동향 및 선진사례를 조사하고 공통된 시사점을 도출하여 연구데이터의 보존 및 활용 모델을 제시한 연구(김선태 외 2010)가 수행되었다.

또한 연구데이터의 장기보존 및 접근, 공유, 재이용의 필수조건으로 디지털 큐레이션 정책의 수립을 강조하면서 외국의 큐레이션 정책 및 서비스 사례를 조사하여 생명주기 단계에 따른 디지털 큐레이션 정책 프레임워크를 제안한 연구(안영희, 박옥화 2010), 미국, 영국, 캐나다, 호주의 15개 정부연구비지원기관에서 제공하는 데이터관리 정책을 분석하여 국내 데이터 정책 개발을 위한 참고사항을 제시한 연구(김지현 2013) 등 보다 정책적인 측면에 초점을 두어 수행된 연구들도 있다.

김은정과 남태우(2012)는 연구자들의 연구 데이터 제출의도를 측정하고, 이러한 제출의도에 영향을 미치는 요인을 분석하여 연구데이터 수집을 활성화할 수 있는 방안을 제시하였으며, 대학 내 연구자들의 연구데이터 관리에 관한

연구(김지현 2012)로 한국연구재단 지원사업의 최종선정과제 연구책임자들 중 대학에 소속되어 있는 연구자들을 대상으로 설문조사를 실시하여 연구데이터 관리 및 공유에서 나타나는 특성을 분석한 연구도 이루어졌다.

외국에서도 대학 소속 직원들을 대상으로 하여 연구데이터의 공유 및 보존 관련 설문조사를 실시한 연구(Diekema, Wesolek and Walters 2014)가 수행되었다. 대학도서관에서 교수 및 연구진들의 데이터 관리를 지원하는데 통찰력을 제공하고자, 교수 및 연구진, 후원 프로그램 지원 사무실 직원(sponsored programs office staff), 기관 리포지토리 사서의 세 그룹을 대상으로 연구데이터의 공유 및 보존에 관한 설문조사를 하였다. 교수 및 연구진이 연구데이터의 계획, 공유, 저장과 관련한 기술을 갖고 있는지, 소속 기관에서 연구데이터 관리 및 공유에 필요한 인프라가 잘 제공되고 있는지 등을 확인한 결과, 교수 및 연구진은 그들의 연구데이터를 공유하고자 하지만 실제 이를 효과적으로 실행하기 위한 기술이 부족했으며, 리포지토리 매니저와 후원 프로그램 지원 사무실에서 필수적 인프라와 지식을 제공함에도 이러한 지원이 그들에게 적절히 작용되고 있지 않은 것으로 나타났다.

연구데이터 리포지토리에 대해 외국에서 진행된 연구들을 살펴보면, 주로 리포지토리 구축, 개발 및 활성화 전략 등을 중심으로 전반적인 개선 방안을 제시한 연구들이 상당 수 진행되어왔다. 연구 대학기반 오픈액세스 리포지토리 구축에 관한 연구(Oppenheim, Probets and Gadd 2003) 및 오픈액세스 리포지토리와 학술저널의 운영모델에 관한 연구(Swan et al. 2005;

Velterop 2005; Swan and Brown 2005)가 이루어졌으며, 여러 학문 분야별 리포지토리의 개발 및 활성화 전략을 위해 연구자들의 요구를 분석한 TARDIS(Targeting Academic Research for Deposit and Disclosure) 프로젝트(Simpson 2005), 사회과학분야 연구데이터의 보존을 위한 리포지토리의 대표 사례로 미국의 ICPSR(Inter-university Consortium for Political and Social Research)과 영국의 UKDA(UK Data Archive) 등을 분석, 디지털 아카이빙 과정 및 장기보존과 활용을 위한 방안을 제시한 연구(Gutmann et al. 2004; Vardigan and Whiteman 2007) 등이 수행되었다.

국내의 경우, 기관 리포지토리의 운영요소, 구축 장애요인 등에 대한 고찰 및 외국 기관 리포지토리의 발전추이와 대표적 사례를 조사하고 국내 구축현황을 살펴본 후, 기관 리포지토리를 구축하고 운영하기 위해 필요한 각 추진주체별 역할에 대해 제안한 연구(황혜경, 김혜선, 최선희 2004), 문헌조사 및 사례연구와 관련 분야 전문가를 대상으로 한 심층면담을 통해 기관 리포지토리 성공을 위한 주요 요인들을 도출하여 오픈액세스 기반 학술정보유통 활성화에 필요한 기초자료를 마련한 연구(황혜경, 이지연 2009), 성공적 리포지토리 구축과 운영 시 유용한 전략적 토대를 마련하고자 세계의 실 운용 중인 리포지토리 중 성공 사례를 조사하여 그 실태와 운영정책을 분석한 연구(정영미, 이상기 2010)가 있다. 이외에도 한국교육학술정보원 (KERIS)의 dCollection을 중심으로 살펴본 연구들(장금연 2004; 김선미, 이나니 2005; 김현희 외 2005; 김현희, 정경희, 김용호 2006)과 IKIS(Institute Knowledge Inventory System)

를 중심으로 운영주체, 참여기관 및 이용자 입장에서의 문제점 및 요구 사항 등을 조사한 연구(김현희 외 2006) 등도 있다.

본 연구에서 중심을 두고 있는 국내의 대표적 인문사회분야 연구데이터 리포지토리인 기초학문자료센터와 관련하여 수행된 연구는 기초학문자료센터 설립에 적절한 제도적·정책적·경제적 운영방안을 제시하는 것에서 시작된다. 이수상 외(2007)는 기초학문자료센터의 오픈액세스시스템 구축 시 정책적으로 고려해야 할 사항과 온라인 커뮤니티 서비스의 주요 기능, 정보공유 활성화를 위한 전략적 과제 등을 제시하여 성공적인 오픈액세스시스템 운영방안을 제안하고자 하였으며, 임건면(2007)은 기초학문자료센터의 설립취지에 부합하는 역할을 충분히 수행을 위해 국내외 학술정보를 제공하는 각 기관들의 주요 사항들에 대한 검토를 통한 기초학문자료센터의 제도적·정책적 개선방안을 제시하였다. 또한, 국가적 차원의 연구데이터에 대한 디지털 아카이빙 시스템 구축이 민간부분에서는 수행할 수는 없는 매우 공공적인 성격의 사업으로 정부가 공공사업으로 추진할 필요가 있음을 제시한 연구(장진규 2007)도 있다.

기초학문자료센터 설립 및 운영 이후 기초학문자료센터가 제공하는 정보서비스의 지식구조와 그 서비스의 이용 과정에서 발생하는 연구자의 정보문제를 분석하고 문제해결과 관련된 정보요구를 충족시키기 위한 지식조직 체계 및 의미기반의 분석적 탐색서비스 모형을 제시한 연구(고영만 2010), 토대기초연구지원사업의 성과확산 강화를 위해 토대기초연구사업이 각 관련 분야 연구의 기초자료로 활용될 수 있는 실제적인 방법으로서 기초학문자료센터와의 연계

방안을 제시한 연구(배경재 2011)로 이어졌으며, 신영란과 정연경(2012)은 기초학문자료센터에 대한 분석과 국내외 아카이브에 대한 사례 조사를 통하여 국내 인문사회 연구데이터 아카이브의 개념적 모델을 제안하였다.

위와 같은 연구들은 연구데이터의 관리 및 공유와 관련한 연구들은 주로 연구자 대상 설문조사 및 선진사례 조사 등을 통해 연구데이터 관리의 정책적인 측면과 리포지토리의 운영방식에 있어 개선방향을 제시하는 것을 중심으로 하고 있다. 반면 인문학 분야 연구데이터의 관리 및 공유에 있어 실제적인 측면에서의 고찰을 통해 연구데이터 수집과 관리 운영체제의 내부적인 관점에서 현황을 다루고 개선방안을 제안한 연구는 찾아보기 어렵다.

### 3. 국내 인문사회 연구데이터 관리의 대표 사례: 기초학문자료센터

#### 3.1 기초학문자료센터의 설립

한국연구재단(National Research Foundation of Korea, 이하 NRF)의 전신인 한국학술진흥재단(Korea Research Foundation, 이하 KRF)은 인문사회(예체능 포함) 연구데이터의 체계적 관리를 통한 연구성과의 확산과 공유를 위해 2005년 기초학문자료센터의 설립을 계획하였다. 이를 위하여 2005년 12월부터 7개월간의 시범사업을 수행하였으며, 2006년 센터 설립을 시작으로 현재까지 운영해 오고 있다.

과거 KRF의 연구비 지원으로 산출된 기초학문분야의 연구성과물 관리의 경우, 연구자가 연

구기간 종료에 따른 연구보고서(최종보고서 또는 결과보고서) 제출로 연구자의 의무를 완료하고, 재단은 연구자가 제출한 최종 성과물 성격의 자료들을 수집·보존하는 방식으로 진행되었다. 하지만 이러한 방식은 최종보고서의 제출 이전 즉, 연구 수행의 각 단계 및 전 기간에 대한 체계적 관리가 불가능하게 하는 방식으로 연구 수행 중 생산된 산출물들에 대한 보존 및 후속 연결을 어렵게 하였다. 따라서 인문사회분야의 연구성과물의 표준화 및 체계화와 다양한 자료들의 다각적 활용을 목적으로 하는 체계구축의 필요성이 대두되었으며 KRF는 연구 지원과 관리의 효율성 제고를 위하여 기초학문자료센터를 설립하게 되었다. 또한 이를 통해 인문사회분야 전반에 걸친 효율적인 연구지원계획을 수립하고, 기초학문자료를 제공함으로써 중복 연구를 방지할 뿐만 아니라 연구가 미비한 분야를 발굴 개발할 수 있을 것이라 기대하였다. 더 나아가 관련 연구 및 후속 연구의 진흥을 도모할 수 있으며 기초학문분야의 발전에 기여할 수 있을 것으로 예상하였다(임건면 2007).

#### 3.2 기초학문자료센터 사업 내용 및 시스템 구성

기초학문자료센터는 크게 '시스템 설치 및 운영'과 '분야별 기초학문자료 DB 구축' 사업으로 나뉘어 진행되고 있으며, 이를 통해 구축된 KRM 시스템은 재단 연구과제의 연구성과물 데이터베이스(이하 DB), 인문사회분야 학술용어 데이터레지스트리(Data Registry: 이하 DR), 그리고 오픈아카이브 시스템으로 구성되어져 있다. 즉 기초학문자료센터는 물리적 공간이 아니며, '서지레코드 기능요건(Functional Requirements

for Bibliographic Records: 이하 FRBR)’을 기반으로 구축한 인문사회 분야 연구성과물 데이터베이스(이하 DB)와 시소러스 기반의 인문사회 분야 학술용어 DR로 구성된 정보시스템의 명칭인 것이다(고영만 2010).

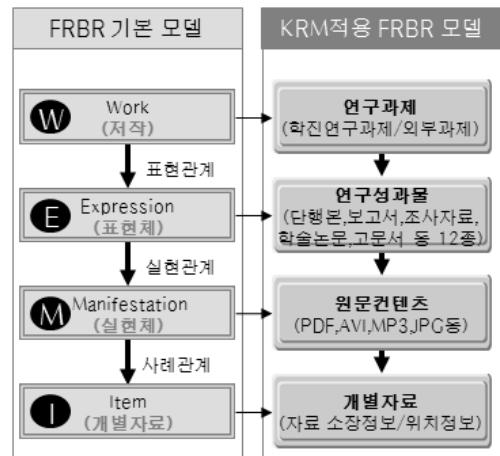
과학 분야뿐만 아니라 인문사회 분야 연구자 또한 하나의 연구 과제를 수행함으로써 다양한 유형의 연구 성과물을 생성해 낼 수 있으며, 연구 수행 과정 중 산출되는 자료의 종류 또한 그 유형과 개수가 다양하게 산출될 수 있다. 이에 기초학문자료센터 시스템에서는 다양한 유형과 매체를 가지는 연구성과물(원자료, 중간산출물, 그리고 연구결과물)의 복잡한 구조를 제대로 기술하고자 계층적 네트워크 구조를 제시하는 FRBR 개념 모형을 연구성과물 DB 구축 시스템에 도입하였으며, 고영만(2010)은 이를 통해 얻고자 하는 목적을 크게 세 가지로 나누었다.

첫째, 논리적 수준으로 계층구조의 목록을 작성함으로써 이용자가 원하는 버전의 자료를 정확하게 선택할 수 있고, 모든 계층의 서지 개체에 접근할 수 있도록 하는 것이다.

둘째, 연결 구조를 지닌 목록을 작성함으로써 그 관계의 정의에 따라 관련 있는 모든 연구 성과물에 대한 검색이 가능하도록 한다는 것이다.

마지막으로, 연계가 가능한 모든 정보로의 연결기능을 강화하고, 검색의 진행과정을 기존의 직선구조에서 네트워크 구조로 변경하여 이용자가 원하는 정보로 바로 이동할 수 있도록 하자 했다는 것이다.

FRBR 개념 모형을 KRM에 적용한 모형은 <그림 1>과 같다.



<그림 1> FRBR 모형을 적용한 KRM 연구성과물 DB 개념도(고영만 2010)

KRM의 연구성과물 DB는 인문사회 학술용어 DR을 통해 언어적 용어관계 모형, 패싯기반 용어의미관계, 범용 온톨로지의 세 모형에 의해 구조화 되어 있으며, 정보이용자는 이 세 개의 지식조직체계 모형이 제공하는 다양한 속성을 이용하여 효율적인 정보 검색을 수행할 수 있도록 하였다. 이 시스템은 연구성과물의 주제색인 어를 토대로 용어 간의 기초적인 의미관계를 표현하고 있어 연구성과물 간 의미적 연관관계의 파악과 검색을 지원하고 있다. 구축기에 등록된 전체 용어 수는 74,573용어이며, 용어 간 관계 설정 수는 168,700여건이다(고영만 2010).

### 3.3 기초학문자료센터 DB 구축 및 이용 현황<sup>1)</sup>

KRF는 기초학문자료센터의 효과적 운영을 위하여 재단이 주관하는 중앙관리센터를 두고 이를 통해 학술활동 결과물 수집·관리, DB 구

1) 한국연구재단. 2015. ‘2014년도 학술자원 공동활용체계 구축사업(기초학문자료센터) 결과보고서’.

축 및 품질관리, 이용자 서비스 개선, 서비스 안정화 등 주도적으로 사업 추진하는 방식을 채택하고 있다(〈그림 2〉 참조).

실질적 DB구축 업무는 중앙관리센터에 의해 선정된 DB 구축 사업단에 의해 수행된다. 사업 초기부터 2009년까지 분야별(7개~9개) 사업 단을 두어 각 분야별 DB의 구축 체계를 마련하였으며, 이후 사업단을 통합하여 2015년도 현재 인문학 분야(예체능포함)와 사회과학 분야 2개의 대표 사업단을 운영하고 있다. 두 사업단의 핵심 업무는 소급자료 수집 및 학술적 전문성을 필수적으로 요하는 원자료와 중간산출물에 대한 DB구축이다.

중앙관리센터의 구체적인 DB 구축 사업 운영 목표는 다음과 같다.

- 재단의 원자료 및 중간산출물 수집을 위한 과제 분석 지원

- 원자료 및 중간산출물 중심의 DB구축
- DB구축사업단 지원기간을 기준 1년에서 2년으로 확대하여 사업단 전문인력의 전문성 제고 및 안정적 사업 운영 도모
- 정부 3.0 정책에 부응하는 인문사회 연구 성과 협력네트워크 구축(50개 기관)
- 정부내 칸막이를 제거하는 타부처 산하 기관 협력 확대
- 국제적 정보유통 지원을 위한 이용자 웹 서비스 영문화 구축
- 특히, 학술적 가치가 높은 기초학문 분야 연구의 학술활동결과물 수집 및 공동 활용
  - DB 643,459건 구축, 1,505여만 건 검색 · 이용('14년말 기준)(〈표 1〉, 〈표 2〉 참조)
  - 토대연구DB(19종)의 온라인 공개서비스 환경 구축



〈그림 2〉 기초학문자료센터 사업 운영 전체 구조

〈표 1〉 기초학문자료센터 DB 구축 현황

(기준: '14.12.)

년도	구분	단행본	보고서	학술 논문	조사 자료	고문서	고도서	이미지	동영상	녹음 자료	웹 사이트	기사 자료	낱장 자료	총건수
2006	자료	361	1,124	4,669	259	454	1,463	567	292	96	6	2,028	827	12,146
	파일	324	1,017	3,475	1,248	81	1,398	4,187	1,167	1,037	3	3,686	5,376	22,999
2007	자료	214	3,134	5,933	306	1,483	2	119	169	68	42	11	1,750	13,231
	파일	142	3,291	5,899	913	3,622	2	2,944	1,861	1,032	1,186	10	11,966	32,868
2008	자료	260	3,998	6,412	319	363	92	706	152	86	7	177	2,062	14,634
	파일	354	4,980	6,504	1,214	1,033	1,098	7,245	2,796	403,319	2	263	6,833	435,641
2009	자료	754	3,899	3,783	619	546	-	277	67	279	22	2	1,233	11,481
	파일	771	5,064	4,111	1,096	442	-	5,844	171	53,002	18	60	27,239	97,818
2010	자료	84	2,074	2,662	254	-	-	68	35	135	0	10	341	5,663
	파일	110	2,367	2,955	615	-	-	3,595	121	808	0	28	632	11,231
2011	자료	230	5,203	4,648	193	13	17	49	17	81	8	2	133	10,594
	파일	0	7,080	92	511	1012	0	264	6	931	0	1	1269	11,166
2012	자료	3,005	9,699	13,562	133	1	20	10	21	32	8	1	189	26,681
	파일	3	13,111	9	447	1	0	308	12	279	0	2	323	14,495
2013	자료	1,265	3,373	2,312	178	1	32	19	74	58	12	0	329	7,653
	파일	0	4,633	10	670	0	0	488	20	137	1	0	567	6,526
2014	자료	1,423	4,406	4,499	101	2	0	17	87	48	34	0	138	10,755
	파일	53	4,385	4082	435	0	0	537	21	876	0	0	326	10,715
계	자료	7,596	36,910	48,480	2,362	2,863	1,626	1,832	914	883	139	2,231	7,002	112,838
	파일	1,757	45,928	27,137	7,149	6,191	2,498	25,412	6,175	461,421	1,210	4,050	54,531	643,459

〈표 2〉 기초학문자료센터 검색/이용 현황

(기준: '14.12., 단위: 명)

년도	2008	2009	2010	2011	2012	2013	2014	합계
이용자수	13,850	1,172,116	1,044,575	1,351,184	1,590,998	3,216,622	6,661,784	15,051,129

### 3.4 기초학문자료센터 DB 구축 사업단

우리나라에 연구데이터의 가치에 대한 인식이 확산된 것은 그리 오래되지 않았다. 과거 개인과 정부 모두 연구데이터의 관리부분에 대한 인식과 수집 체계가 부실하였고, 이로 인해 오랜 기간 동안 엄청난 양의 연구데이터가 관리되지 않은 상태로 방치되고 있었다. 연구데이터의 유실은 연구자들이 선행적으로 유사 연구가 존재했음에도 불구하고 후속 연구 수행에 있어서 선행과 마찬가지의 관련 데이터들을 수집하는

것부터 연구를 시작해야하는 비효율적·비효용적 행태를 빈번히 야기 시켰다. 최근 들어 연구데이터의 가치에 대한 인식과 체계적 수집·관리의 필요에 대한 목소리가 높아지면서 연구데이터의 학술적·경제적 가치 그리고 국가자원으로서의 활용성 측면의 가치까지 그 중요성이 확대되었다. 이는 이길우(2007)가 주장하는 바와 같이 근본적으로 연구성과는 확산과 활용 및 상용화를 통하여 부가가치가 실현되기 때문이다.

KRF는 기초학문자료센터를 통해 연구데이터 활용성 극대화를 위한 아카이브를 구축하고

자 하였으며, 데이터 아카이브를 통해 활용 및 보존의 가치가 있는 연구데이터들을 발굴하여 수집, 정리, 보존하는 전문적서비스를 제공함으로써 연구와 교육 등에서 다면적 활용이 가능하도록 하고 있다. 데이터 아카이브의 기본 조건은 아카이빙 대상 자료가 디지털 정보자원이어야 한다는 것이다. 하지만 세상에 존재하는 모든 자료가 태생적 디지털 자료일 수는 없어 아날로그 자료를 디지털화하는 작업이 요구 된다. 따라서 기초학문자료센터는 KRM DB 구축 사업단을 따로 두어 자료의 체계적인 수집·정리의 업무를 수행하게 하고 있다. 사업단의 핵심적 역할은 학술연구조성사업에 의해 지원된 연구과제의 연구성과물 수집을 시작으로 수집된 연구성과물 원문 자료의 디지타이징 및 메타데이터 작성과 검증, 해당 주제분야 용어DR DB 구축, 그리고 수집 자료에 대한 저작권 동의서 확보까지의 업무 수행이다.

기초학문자료센터의 수집 대상인 연구성과물은 연구결과물을 비롯하여 연구를 수행하면서 생성된 관련 원자료와 중간산출물 등이 모두 포함된다. '원자료'와 '중간산출물'의 용어 정의는 다음과 같다(임건면 외 2007).

- 원자료: 연구자들이 연구과제를 수행하는 과정에서 수집·참고한 일체의 자료  
(예: 고문서, 논문, 보고서, 단행본 기사, 유인물, 일기, 사진자료, 음성자료, 동영상자료, 통계자료 등)
- 중간산출물: 연구과제를 수행하는 과정에서 작성한 일체의 자료 중 최종연구보고서를 제외한 것

이들 자료들은 모두 기초학문자료로써 관련

분야 후속 연구에서 지속적으로 활용될 수 있는 원자료적 가치를 지닌 것으로 특정 학문분야에 국한되어 생성되는 것이 아닌 다양한 학문 분야에서 연구의 성격과 방법에 따라 언제든지 신출 가능한 양적·질적 자료들이다. 물론 분야별로 데이터에 대한 개념과 유형에 차이를 보이고 데이터 관리가 연구과정의 일부로서 해당 분야의 고유한 특성을 나타내고 있다고는(Beagrie et al. 2009; Borgman 2012; 김지현 2012) 하지만, 과학분야에서의 신규 데이터 생성이나 인문사회과학분야에서의 활용이나 정도의 차이 일 뿐 거의 모든 학문분야에서 데이터를 이용하여 연구문제를 도출하고 이를 이용하는 연구방법을 혼용하는 등 연구과정 전 기간에 걸쳐 발생된 연구데이터인 것은 분명하다.

#### 4. 기초학문자료센터를 통한 인문학 분야 DB 구축

연구자의 연구행태는 학문 분야와 시간의 흐름에 따라 상이하다. 과거로부터 많은 인문학자들이 질적자료에 대한 이차적 해석과 분석을 통한 연구 수행의 행태를 보여 오고 있는데 이는 자연과학자들 심지어 사회과학자들과의 절대적 차이이기도 하다. 인문학자들의 질적자료에 대한 의존은 연구에 있어 핵심적 연구 대상이 단행본, 고문서, 기록물 혹은 실존 역사물 등의 인쇄형 자료와 비인쇄형 자료인 것으로 나타났으며, 현재까지도 이들 자료에 대한 활용은 제고의 여지가 없다. 하지만 시간의 변화와 기술의 발전은 이들 자료의 접근 방식과 보존형태의 변화를 가져왔다. 단행본이나 고문서, 기록물 같

은 인쇄형 자료뿐만 아니라 실사물의 경우에도 이미지나 음성, 영상 등의 파일형태로 변환·저장하여 디지털 정보자원으로 유통되고 있다. 이는 이들 자료의 손망실을 최소화하고 영구적 보존을 위한 노력이다. 따라서 인문학분야에서도 젊은 연구자들이나 인터넷 사용에 무리가 없는 연구자들로 하여금 간편하게 웹상에서 정보 또는 디지털화된 데이터 손쉽게 획득하여 활용하게 하는 행태로 정보행위 자체의 변화를 이끌고 있다. 하지만 아직도 많은 인문학자들이 디지털 방식과 기존의 아날로그 방식을 혼용하여 사용하거나 아날로그적 연구행태를 고수한다는 사실을 간과해서는 안 된다. 인문학 또한 연구 데이터 중심의 연구라는 사실에 동의하기는 하나 방대한 양의 산발적으로 흩어져있는 아날로그 방식의 대상 자료들을 수집·정리·보존하고 전문적 서비스를 통해 가치를 부여함으로써 그 활용적 가치를 증대시키는 일은 그리 간단하거나 쉬운 일은 아니다.

따라서 본 장에서는 실제 수행 중인 기초학문 자료센터의 인문학 DB 구축 사업의 진행 과정과 사례를 살펴봄으로써 인문학 분야 연구데이터 아카이브 구축의 현 상태와 발전 가능성은 짚어보고자 한다.

#### 4.1 인문학 데이터의 수집

기초학문자료센터는 KRM DB 구축을 위한 중앙관리센터를 설치해 자료 수집 업무를 담당하고 있다. 단, 년간 처리해야하는 과제 수 대비 담당업무를 수행할 자체 인프라 체계구축의 미비로 외부 DB 구축 사업단을 선정하여 수집 준비에서 DB 구축까지의 업무를 위임하고 있다.

자료의 수집·보존·활용·보급 등에 있어 가장 큰 문제는 표준의 제시와 법적 문제의 해결이다. 따라서 중앙관리센터는 시범구축과 여러 건의 정책연구를 통해 KRM DB 구축을 위한 조직과 구성, 시스템 기능 점검, 메타데이터 표준안 마련, 법적 처리 지원 등의 기반을 마련하여 중앙관리를 하고 있다. 본 논고에서는 인문학 분야의 연구데이터 관리에 대해 논하기 위하여 인문학과 사회과학으로 나누어진 DB 구축 업무 중 인문학 분야 DB 구축 업무에 대해 보다 상세히 다룬다.

인문학은 인간과 인류문화에 관한 학문으로 과거로부터 현재, 그리고 미래의 변화와 가치 창출의 예측까지 연계되어 있는 학문분야라 할 수 있다. 이러한 측면에서 인문학 분야 데이터의 보존과 유통 및 재이용은 인문학 분야의 지속적 발전과 나아가 고부가가치 창출의 지식산업으로의 육성에 있어 매우 중요한 역할을 담당할 것이다. 김기덕, 이상훈(2002)은 인문학 자료들의 수집 범위에 대하여 폭넓은 시각의 필요를 언급하였다. 인문학 자료에 대한 평가는 시대에 따라 변할 수 있으며, 따라서 현재의 기준만으로 자료의 가치를 평가하는 것은 후세에 높은 평가를 받을 가능성이 있는 자료에 대한 합당한 조치라 할 수 없다 하였다. 즉 현재의 자료적 가치에 대한 평가와는 별도로, 시대와 세태 등을 상징적으로 나타내는 인문학 자료들을 폭넓게 수집·보존한다는 것은 미래적 입장에서도 역사적·문화적·사회적으로 큰 의의를 지니는 것이다.

따라서 인문학 분야 데이터 혹은 연구데이터들은 다른 어떤 분야의 데이터들보다 장기보존적 또는 항구적 가치를 지닌다 할 수 있으며 이

들 데이터의 체계적 관리는 국가적·개인적 의미가 되어야 하는 것이다. Beagrie와 Greenstein (1998)은 장기보존과 활용의 목적으로 연구데이터를 전문적으로 관리하는 아카이브를 학술적 데이터 아카이브라 하였으며, 신영란 외(2012)는 활용 및 보존의 가치가 있는 연구데이터를 발굴하여 수집, 보존하고 연구와 교육에 활용할 수 있도록 전문적인 서비스를 제공하는 기관이 데이터 아카이브라 하였다. 즉, 데이터 아카이브는 연구데이터를 수집·정리·보존하는 것 뿐만 아니라 공유와 재이용이 가능하도록 제공·제공하여 연구와 교육의 활성화를 지원하는 서비스 기관인 것이다. 이러한 의미에서 기초학문자료센터는 매우 선진적 데이터 아카이브라 할 수 있는데, 이는 기초학문자료센터의 구체적 목표<sup>2)</sup>가 다음과 같기 때문이다.

첫째, 학술연구조성사업에 의해 지원된 연구 성과물 DB 구축

둘째, 국가적 차원의 기초학문분야 통합자료 센터 및 아카이빙 시스템 구축

셋째, 구축 자료의 관리, 공유, 확산을 위한 기반체계 개발

기초학문자료센터의 인문학 아카이브 구축 대상은 인문학 분야의 종교, 어문, 철학, 역사 등 과 관련한 21개 주제 분야 과제들이다. 주요 수집 대상은 연구데이터로써 연구 전 과정에서 산

출된 유의미한 가치를 지닌 데이터들로 자료의 유형은 크게 텍스트, 이미지, 음성, 동영상 등으로 나누어진다. 따라서 인문학 분야 DB 구축팀은 연구데이터의 수집에 있어 인문학적 이해를 바탕으로 한 기본원칙과 프로세스를 마련하여 각각의 특성을 지닌 과제들의 처리 효율을 높이고자 하였다.

기초학문자료센터 인문학 분야 DB 구축은 다음과 같은 몇 가지 핵심적 단계를 거쳐 이루어진다.

#### ① 과제 배분

: 중앙관리센터는 해당년도에 구축되어야 하는 연구과제들의 목록을 사업단 별로 배포

#### ② 과제 분석<sup>3)</sup>

: 각 사업단은 배당 받은 연구과제들에 대한 기본정보 분석을 시작으로 각 연구과제의 원자료, 중간산출물, 연구결과물 등의 연구 성과물 수집을 위한 추정작업 실시(분야별 주제 전문가 참여)

#### ③ 과제분석지 전달

: 각 사업단은 과제 분석을 완료한 분석지를 정리·검토하여 KRM 중앙관리센터로 송부하고 이는 다시 해당 연구자들에게 전달

#### ④ 과제 분석 피드백

: 전달된 과제 분석지의 내용을 검토한 연구자들이 해당 과제의 연구데이터 제공(저작권 동의서 함께 제출) 또는 과제분석지에

2) 2014년도 기초학문자료센터 지원 사업 결과보고서.

3) 과제분석이란 한국연구재단의 학술연구조성사업에 의해 지원된 연구성과물 DB(기초학문자료센터) 구축을 위해 연구자가 제출한 연구결과보고서 및 연구결과물의 분석을 통하여 연구에 사용된 원자료의 실체와 중간출물의 유무(수량, 종류, 내용 등 포함) 등을 추정하는 작업이다. 과제분석을 통해 연구자가 제출한 연구성과물 이외의 원자료 또는 중간산출물 등의 추가 존재가 추정된다면, 이러한 추정을 근거로 중앙관리센터는 연구자에게 보다 구체적인 자료협조 요청을 할 수 있다. 따라서 과제분석은 그 자체에 목적이 있기 보다는 궁극적으로 연구자의 자발적인 자료기탁을 원활하게 하려는데 의의가 있는 것이다. 또한 DB 구축에 있어 구축업무 담당자는 과제연구자의 검토와 의견이 첨부된 과제분석 결과를 바탕으로 작업을 수행함으로써 작업의 효율과 DB의 품질 향상을 도모할 수 있다.

### 대한 검토 의견 제시

#### ⑤ DB 구축

: 각 사업단은 피드백 내용에 따라 또는 자체 검토 결과에 따라 DB 구축 대상을 구분하고 과제별로 메타데이터 입력 및 해당 자료를 업로드 실시

#### ⑥ 검증

: 해당 연도에 구축한 DB 내역에 대한 다면적 검증을 실시하고 오류사항에 대한 수정·보완

#### ⑦ 기구축 DB 보완

: 해당 연도의 연구과제 구축을 중심으로 하되 기 구축 된 DB 내역 중 발견되는 오류 사항에 대해서는 정리하여 즉각적인 수정·보완을 실시하고, 연구자의 협의 또는 동의를 거쳐야하는 사안에 대해서는 중앙 관리센터에 내용을 고지, 이와 더불어 수정·보완 내역에 대한 재검증 실시

위에서 제시한 데이터의 수집과 보존 처리를 위한 과정에서 가장 중요한 부분은 과제분석이다. 과제분석은 연구자가 이미 수행 완료한 연구과제를 대상으로 연구과제의 요점을 요약하고 원자료, 중간산출물, 연구결과물을 명확하게 선별하여 추출하는 것을 핵심으로 연구자가 연구 종료 후 제출한 자료 이외의 연구데이터의 수집 가능성을 넓히고자 실시하는 것이다. 따라서 과제분석 결과에 따라 KRM DB의 질이 결정될 수 있으므로 반드시 연구과제에 대한 충분한 이해가 가능한 유사 주제분야의 전문가에 의해 실시되어야 하며, 소수의 책임편집인에 의해 검토 및 수정 과정을 거친 후 연구자에게 전달되어야 한다. 또한 과제분석 전 주제분야별 샘플

분석을 통해 주제분야의 특수성을 포괄 할 수 있는 상세한 가이드라인이 마련되어야 한다.

한 과제당 과제분석은 1종이 작성된다. 과제 분석 1종은 1) 기본 과제 정보 및 연구의 개요, 2) 중간산출물 추정, 3) 원자료 추정, 4) 연구 결과물 추정의 네 부분으로 구분되어 구성 된다. 과제 분석의 목적이 해당 과제의 원자료 및 중간산출물 유무 및 자료의 성격을 파악하는 것이므로 연구의 내용, 자료의 유무에 따라 과제 분석의 깊이와 보고서 길이는 매우 다양하게 나타난다. 예를 들어 과제 분석자가 특정 과제에서 원자료나 중간산출물을 기대하기 힘들다고 파악한다면 과제의 내용에 대한 간략한 요약으로 과제 분석을 갈음하지만, 다양한 자료가 존재하고 자료의 성격에 대한 상세한 기술이 가능하면 과제 분석 내용도 복잡한 성격을 띠게 된다. 고영만(2010)은 과제 분석의 연구자 전달과 이에 대한 연구자 의견 제공은 자료에 대한 전문적 이해를 바탕으로 자료의 위상, 내용, 특징 및 타 자료와의 관련성을 종합적으로 파악하여 질적 수준이 강화된 자료를 구축하게 하며, 이와 더불어 전문적 지식에 기초한 주제분류/주제용어/용어 DR 등 메타데이터를 작성하고 자료 간 관계 맺기를 통해 선행/후속 연구의 여부, 연관 관계의 여부 등의 점검이 가능하게 한다고 하였다.

## 4.2 인문학 데이터의 공유와 재이용

학술 커뮤니케이션 관점에서 연구자의 연구 수행은 학술적 이해형성이라 말할 수 있다. 학술적 이해형성은 연구자가 연구문제 해결을 위해 경쟁적 관계를 갖는 다른 연구자와 연구 성

과를 지속적으로 공유함으로서 문제에 대한 해결과 이해의 명백성, 적합성, 타당성을 증진해 나가는 정보행위라 할 수 있다. 연구자는 자신의 가설을 뒷받침하고 주장의 타당성과 완성도 확보를 위해 적합하고 신뢰할 수 있는 다양한 성격의 연구문헌과 그 연구문헌에 관한 데이터를 필요로 한다. 즉 연구자는 학술커뮤니케이션의 담화 네트워크상에서 해당 연구의 주제영역에 관한 기존의 연구문헌 간에 성립하는 관계들을 파악하고, 자신의 정보문제와 정보요구에 적합한 정보를 검색하고 활용하는 등의 정보행위를 하게 된다(고영만 2010). 따라서 연구데이터를 활용한 연구에 있어 풍부하고 상세한 메타데이터를 제공하는 서비스를 이용함으로써 활용하는 연구데이터의 전체적인 맥락적 이해가 충분히 가능하도록 하여야 하는 것이다. Pike and Gahegan(2007)은 학술적 이해형성의 정보 행위를 3단계: 연구주제 및 아이디어의 발굴과 수집단계, 연구수행 단계, 연구결과에 대한 구조화된 정보자료 작성 및 배포로 구분하였으며, 고영만(2010)은 Buckingham(2006)이 제시한 연구수행 과정에서 단계별로 이루어지는 학술적 이해형성을 위한 연구문헌의 활용목적 5유형을 응용하여 KRM의 지원 기능을 여섯 측면에서 비교 분석한바 있다. 이러한 학술 커뮤니케이션에서의 학술적 이해형성 또한 자원 공유 문화가 안정화된 환경에서 전문적 서비스 지원 시스템을 통해 활성화 될 수 있다.

디지털 정보 자원으로서의 연구데이터의 공유와 재이용에 관심은 데이터 큐레이션과 생애주기 모델에 대한 관심으로 이어졌다. 데이터 큐레이션은 연구성과물들을 체계적으로 관리하는 과정에서 데이터의 사용을 장려하고 관리하

는 활동과 데이터 검색이 편리하도록 지원하기 때문이다. 또한 데이터 생애주기 모델에 대한 이해를 통해 연구데이터 공유와 재이용에 있어 미래가치를 창출할 수 있기 때문이기도 하다. 하지만 우리나라의 경우 아직까지 김지현(2012)의 연구 결과처럼 연구데이터 공유와 재이용에 대한 인식수준은 낮은 편이다. 그의 연구에 따르면 연구데이터의 공유와 재이용은 학문발전의 촉진이라는 궁극적인 효과와 타인의 연구에 기여한다는 점, 공유된 데이터에 바탕을 둔 응답자 본인의 논문 인용 수가 증가할 것이라는 긍정적 기대를 지님에도 불구하고, 데이터의 표절의혹, 부정확한 데이터 분석, 불필요한 시간과 노력의 소요 등의 장애요인으로 공유를 꺼려한다고 밝히고 있다. 또한 연구자들은 데이터 접근/이용권한의 본인 결정 여부, 저자 승인에 의한 배포, 정부지원 과제의 공공재로서의 공유 등을 연구데이터 공유 조건 요구한다고 언급하였다. 같은 맥락에서 연구데이터의 공유와 재이용 또한 기초학문자료센터가 설립시기부터 지금까지 계속 고민해오고 있는 부분이다. 이에 대한 우선적 해결을 위해 자료 작성자 스스로 데이터 구축을 진행하는 납본 유형의 셀프아카이빙 시스템을 제공하고 있으며, DB 구축 사업단의 업무 지원을 위하여 저작권 동의서를 연구자들에게 함께 제출받고 있다. 이외에도 자료 기탁 시 공개 여부 또한 제공자의 의견에 따라 처리되도록 하고 있다.

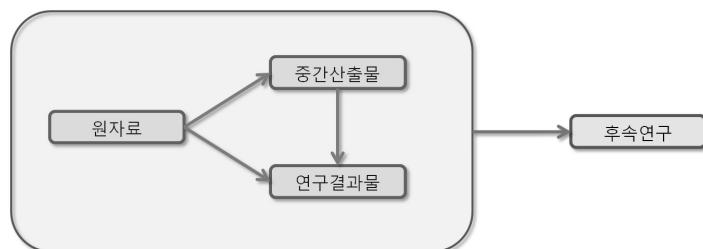
인문학 분야 연구과제의 대부분은 타 학문분야에 비해 상대적으로 중간산출물이 산출되는 비율이 낮다고 볼 수 있다. 반면에 대부분의 연구가 연구 수행 이전에 존재한 다양한 텍스트 및 매체 자료를 면밀하게 분석하고 비교하는 등

의 원자료 중심 연구형태로 수행되고 있는 것이 특징이다. 따라서 인문학 DB 구축의 활용성 증대를 위해 연구데이터에 대한 큐레이션이 방안이 될 수 있을 것이다. 큐레이션을 통해 인문학 분야의 특성, 즉 연구 주제뿐 아니라 연구 과정 중에 발생하는 학술 커뮤니케이션의 특성적 차이에 대한 이해를 보다 체계화 할 수 있을 것이다. KRM에서는 데이터큐레이션의 초기적 모델로 디지털 지식지도 개념을 도입하여 2011년부터 시범 구축하여 그 체계를 마련해 가고 있다.

기초학문자료센터 DB 사업 내에서 디지털 지식지도란 연구의 순환과정에 따른 자료의 활용, 생성을 보여주기 위해 시범적으로 구축되는 DB 구축 모형이다. 문학, 사학, 철학 등 대부분의 인문학 연구에서는 기존의 작품이나 고전 텍스트를 읽고 해석하는 연구가 주류를 이루며, 이들 자료의 일부는 이미 단행본이나 온라인 텍스트로 존재하고 있어 저작권의 문제로 인해 원문을 DB로 구축하는 것은 어렵다. 하지만 기초학문자료센터 DB가 원자료 - 중간산출물 - 연구결과물이라는 연구의 순환과정에 따른 자료의 활용, 생성을 보여주기 위해 중간산출물과 연구결과물 이외에 원자료까지 DB화하여 연구자나 학습자들로 하여금 활용하게 한다면, 후속연구자는 선행연구에 대한 탐색에 있어서 그 전체적

흐름을 더욱 쉽게 이해할 수 있을 것이며, 새로운 연구를 위한 원자료를 수집하는 과정에서 시간과 비용적 측면에서 효율을 얻게 될 것이다. 따라서 디지털지식지도는 연구와 최근접성을 가진 원자료와 연구과정에서 산출된 중간산출물, 연구결과를 집합한 연구결과물의 순환/연결과정을 맥락정보라는 매개를 통해 분석적으로 표현하기 위한 개념 및 실제적인 자료구축 모델이라 할 수 있다.

디지털 지식지도는 실제 활용 가능한 자료를 구축하는 것을 목적으로 하지 않는다. 단지, 해당 연구내 원자료-중간산출물-연구결과물이라는 연구의 순환과정을 맥락정보를 통해 표현해 줌으로써 연구의 전체 구조파악 및 연구내의 자료 활용성을 보다 쉽게 이해할 수 있도록 한다. 하지만 DB구축에 있어 여러 다른 연구에서 동일한 원자료를 사용한 경우 각 연구마다 그 맥락정보가 달라 각 연구마다 원자료에 대한 표현체를 각각 작성해야 하는 문제점이 있다. 이는 한 자료에 대한 표현체의 다수 존재는 향후 DB 관리 및 이용자 서비스 상황에서 혼란을 초래할 가능성이 있음을 의미한다. 따라서 현재 구현되고 있는 디지털 지식지도 모형은 매우 기초적 단계의 것으로 향후 추가적인 연구 및 실제적인 적용에 대한 보다 체계적인 접근이 필요하다.



〈그림 3〉 자료의 순환 및 후속 연계

## 5. 연구데이터 수집 및 활용성 증진을 위한 제안

2012년 6월 영국의 The Royal Society는 “Science as an open enterprise”라는 의미심장한 보고서를 출판했다. 이 보고서는 연구에 있어서 공개의 원칙을 보존하고, 새로운 과학혁명을 창출할 수 있는 방향으로 데이터를 활용하기 위해서는 정보기술로 인해 쏟아지는 데이터 문제를 해결해야 한다고 촉구한다. 보고서의 핵심은 연구자들이 데이터를 표준적인 방법으로 공개 및 공유하고 이를 위한 지원 시스템이 구축되어야 한다는 것이다. 보고서의 핵심은 데이터의 공개와 공유는 연구자 개인, 기관 혹은 국가 연구비 지원기관의 연구성과를 넘어서는 과학 발전의 토대라는 것이다. 분야별, 개인별 차이는 있지만 우리의 연구 환경과 연구자의 인식은 아직 이러한 이상에는 미치지 못한다. 하지만 앞에서 언급한 바와 같은 연구데이터의 중요성 및 공유의 필요성에 대한 관심과 인식 수준은 최근 제고되고 있다. 이 장에서는 향후 국내 학술커뮤니티에서 연구데이터가 보다 효율적으로 수집, 관리되고 데이터의 활용성이 제고될 수 있는 방안에 대해 논하고자 한다. 특히 국가연구비지원을 받아 수행된 연구에서 수집, 생산되는 데이터에 초점을 맞추고자 하였다.

### 5.1 연구데이터의 수집과 공개 관련 정책의 시행

정보서비스에 있어 문현정보학의 주요 관심사 중의 하나는 학술연구에서 생산된 다양한 성과물, 특히 저서, 학술논문을 어떻게 하면 체계적으로 수집, 기술하고 원문에 대한 접근을 제공

할 것인지에 대한 것이다. 이러한 종류의 정보 유통은 비교적 표준화되고 정형화되어 효율적으로 진행되고 있다. 반면 연구데이터의 경우는 연구성과물로 인정받지 못하고, 내용의 다양성, 형태의 비정형성 등으로 인해 상대적으로 관심을 덜 받아왔다. 하지만 2000년대 들어 전자학술논문의 자유로운 접근을 핵심으로 하는 오픈액세스 운동이 활발하게 전개되고 “OECD Principles and Guidelines for Access to Research Data from Public Funding”(Pilat and Fukasaku 2007)과 같은 보고서가 출간되면서 연구자 및 연구비 지원기관의 관심을 끌게 되었다. 공적재원의 지원을 통해 생산된 연구데이터의 공개 접근은 동시에 정부로부터 생산되는 정보와 데이터의 공개접근, 즉 오픈 정부라는 흐름과 맥을 같이 한다.

이러한 배경에서 2011년 미국 NSF는 DMP를 마련하여 의무적으로 제출하도록 규정을 제정하였으며, 같은 해 영국의 RCUK(Research Councils UK) 또한 데이터 정책에 대한 공동 원칙을 발표하고 산하 일곱 개의 연구비 지원기관(Research Council)이 DMP를 포함한 다양한 연구데이터 관련 정책을 시행하게 되면서 (김지현 2013) 연구데이터의 관리와 공유 문제가 개별 연구자뿐만 아니라 대학과 연구소 같은 연구기관에서도 시급하게 해결해야 하는 이슈로 등장하게 되었다. 그러나 연구계획서에 DMP를 의무적으로 제출하여 심사반도록 한 정책은 연구자들에게 상당한 부담과 혼란을 초래하였으며, 이로 인해 미국과 영국의 상당수 대학에서는 대학도서관이 이 분야에 가장 전문성이 있어 지원을 할 수 있는 기관으로 부상하게 되었다.

연구비 지원기관의 DMP 정책은 분명 연구자의 행동을 변화시킬 수 있는 매우 효과적인 수단임에는 분명하다. 하지만 단순히 DMP 정책을 시행하는 것으로 기존의 관행을 모두 변화시키고 보다 효과적인 연구데이터의 관리 및 공개가 실현된다고 할 수는 없다. 이미 DMP 정책을 시행하고 있는 NSF와 영국의 연구비 지원기관으로부터 실제로 DMP가 어떻게 평가되고 있고, 구체적인 영향이 어떠한지에 대해 알려진 바가 없다. 영국의 EPSRC(Engineering and Physical Sciences Research Council)의 경우 다른 연구비 지원기관과 달리 의도적으로 DMP 정책을 시행하고 있지 않고 있다. DMP 정책은 연구데이터의 관리 및 공유의 책임을 연구자 개인에게 부담시킨다는 측면이 강하다. EPSRC의 경우 개별 연구자가 아닌 연구기관 즉 대학에 연구데이터의 관리 및 공유의 책임을 묻고 있다. 한국연구재단의 경우 연구비 지원을 받아 수행된 연구에서 생산되는 데이터와 결과물을 재단에 기탁해야 한다는 규정(인문사회분야 학술연구지원사업 처리규정 등)을 이미 시행 중이다. 하지만 실제로 연구가 완료된 후에 연구성과물이 아닌 연구에서 수집, 생산된 데이터가 존재 및 기탁 여부를 확인하고 이를 지키지 않은 것에 대한 처리 등의 후속 관리에 대한 체계가 마련되어 있지 않다. 특히 데이터의 체계적인 관리에 대한 부분을 규정하지 않고 연구 종료 후에 해당 데이터의 기탁만을 요구하고 있는 것은 그 효율을 더욱 반감시킨다.

데이터관리계획 제출 정책은 연구의 계획 단계에서 연구자로 하여금 어떤 데이터를 수집하

고, 어떻게 관리하고 공유할지 본인이 직접 기술하게 함으로써 향후 이런 절차들이 제대로 수행되었는지 보다 효율적으로 관리할 수 있는 근거를 마련한다는 점에서 보다 효과적인 정책이라고 할 수 있다. 대개의 경우 연구자는 자신의 데이터를 공신력 있는 분야별 혹은 기관 데이터 센터나 리포지토리에 기탁함으로써 데이터 관리계획에서 요구하는 요건을 충족시킬 수 있다. 데이터관리계획은 통상 1~2쪽 내외의 간략적인 내용으로 통상적으로 연구계획서의 본문이 아닌 첨부 혹은 기술문서로 요구되고 있다. 김지현(2013)은 미국, 영국, 캐나다와 호주의 연구비 지원기관 15개를 대상으로 데이터관리 계획을 중심으로 한 연구데이터 관리정책을 비교, 분석한 바 있다. 영국의 Medical Research Council(MRC)<sup>4)</sup>에 따르면 데이터관리계획은 연구 프로젝트 전체를 데이터의 관점에서 정리하는 것이다. 즉 연구의 계획 단계에서부터 연구의 종료에 이르기까지 데이터의 생성뿐 아니라 공유, 보존이라는 데이터 중심의 체계적인 정리 및 계획을 수립하는 것이 데이터관리계획의 목적이다. 학술커뮤니케이션의 관점에서 데이터관리계획을 시행하는 목적은 연구 데이터를 손쉽게 발견하고 접근할 수 있고, 이해하고 평가할 수 있고, 원래 만들어진 목적 이외에 활용할 수 있으며, 상호호환이 가능할 수 있도록 하는 것이다.

데이터관리계획의 내용이 대개 2페이지 내외의 짧은 내용이기 때문에 연구자에게 큰 부담이 되지 않을 것이라는 근거로 연구비 지원기관에서 데이터관리계획을 연구 제안서의 의무사항

4) <http://www.mrc.ac.uk/documents/pdf/data-management-plans-guidance-for-reviewers/>

으로 하는 것은 어려운 일이 아닐 수 있지만 실제로 건전한 연구데이터의 관리 및 공유 환경이 조성되려면 정부 부처, 대학, 학계 및 연구소 등에서 상당한 준비와 지원이 필요할 것으로 보인다. 연구자들이 생산, 수집한 데이터를 기탁할 수 있는 국가 차원의 데이터 센터, 학문 분야별 안정적인 데이터 리포지토리 인프라의 구축이 미약한 상황이다. 이런 측면에서 보면 한국연구재단의 기초학문자료센터는 비록 인문사회 분야에 한정되어 있지만 비교적 안정적인 데이터 리포지토리로서 이미 기능하고 있다는 점에서 예외적인 상황이다. 따라서 이런 토대 위에 체계적인 데이터관리계획을 정책으로 마련한다면 동분야 연구데이터의 수집과 공유에 보다 효과적인 결과를 가져올 수 있을 것으로 기대된다.

따라서 국내의 대표적인 국가연구비 지원기관인 한국연구재단은 해외 연구비지원기관에서 이미 시행 중인 데이터관리계획 정책을 벤치마킹하여 국내에 적용할 방안을 찾는 것이 필요하다. 물론 이에 앞서 인문학 분야적 특성과 인문학 연구자들의 학술커뮤니케이션 패턴에 대한 상세한 이해가 기반 되어져야 한다. 다시 말해 어떤 범위의 연구에 대해 또 어떤 종류의 연구데이터를 대상으로 데이터관리계획을 적용할 것인지를 고려할 필요가 있으며 분야별 특성을 고려한 탄력적인 적용 또한 고민할 필요가 있다.

## 5.2 데이터 관리, 수집 및 공유를 위한 인프라 및 환경 구축

데이터관리계획과 같은 정책이 실제로 시행된다 하더라도 이것이 연구데이터의 관리 및 공

유의 문제를 모두 해결해 주지는 않을 것이다. 궁극적으로는 개별 연구자의 관점에서 연구데이터의 체계적인 관리와 공유가 정책적인 의무 사항이 아닌 정상적인 연구수행의 일부로 인식되고 이에 따른 행동으로 나타나야 한다. 연구자라면 연구를 수행하고 연구의 결과를 저서, 학술대회 논문, 학술지 논문, 특히 등으로 공개하고 출판하는 것을 당연하게 생각한다. 이와 더불어 연구를 위해 혹은 연구의 수행 과정을 통해 수집, 생산된 데이터를 해당 학문 분야에서 인정하는 수준에서, 표준적인 방식을 통해 관리하고 개인정보의 노출, 잠재적으로 부정적인 영향 초래 등과 같은 특별한 사유가 없는 한 공개하는 것은 연구자의 당연한 의무라고 할 수 있다. 하지만 아직 이러한 인식이 많은 연구자에 의해 공유되고 있지 않다. 연구데이터의 필요에 대한 요구는 증가하고 있는 반면 공유의 의식은 미미한 수준이다. 오히려 연구데이터가 연구자 본인의 소유라는 인식과 함께, 데이터의 공유가 연구자의 경쟁력에 저하가 된다는 인식도 있다. 따라서 정상적인 연구의 일부로서 데이터의 관리 및 공유라는 점에 대한 지속적인 인식 개선 노력이 정책 당국 및 학술 커뮤니티에 의해 우선적으로 선행되어져야 한다.

데이터의 체계적인 관리 및 공유는 연구자에게 추가적인 업무의 과중을 가져온다. 그러므로 이러한 업무를 지원할 수 있는 다양한 교육, 서비스의 제공이 필요하다. 이러한 지원체계가 없는 상태에서 단지 데이터관리계획이라는 정책만을 시행한다면 연구데이터에 대한 긍정적인 인식 개선뿐 아니라 효과적인 데이터의 수집 및 공유를 이룰 수 없을 것이다. 따라서 연구비 지원기관, 대학을 비롯한 연구기관, 국가 및

대학도서관을 비롯한 다양한 기관에서 연구자를 대상으로 연구데이터 관련 지원 서비스를 지원할 필요가 있다. 특히 연구자가 소속된 대학 차원에서의 데이터 관련 교육 및 지원 서비스의 제공은 가장 큰 영향력을 가질 것으로 예상된다.

대부분의 연구자들이 전문가로서 직접 데이터 수집계획을 세우고 데이터를 수집하여 분석 하지만, 데이터를 체계적이고 표준적인 절차를 통해 관리하고 공유하는 것에는 전문적이지 않을 수 있다. 연구데이터가 의미를 가지고 재이용되려면 데이터를 기술하는 메타데이터의 작성이 매우 필수적이므로, 표준적인 메타데이터 요소에 풍부한 정보를 기록으로 남기는 것은 전문적인 지식을 필요로 한다. 즉 연구자들이 데이터를 체계적으로 관리하고 공유할 수 있도록 전문적으로 지원하는 인력의 양성 및 활용이 매우 시급한 상황이다. 이러한 이유로 앞서 언급한 The Royal Society 보고서에서도 요구되는 여섯 가지 변화 가운데 연구데이터의 활용을 관리하고 지원할 수 있는 데이터 사이언티스트(data scientists, 이하 데이터전문가)의 양성을 하나로 제시한다. 향후 데이터 아카이브에서 연구데이터와 관련하여 데이터전문가의 필요성 혹은 양성 방안에 대한 논의는 더 큰 비중을 차지하게 될 것이다. 국내의 경우 매우 예외적으로 인문사회 분야를 지원하는 기초학문자료센터와 같은 시스템 환경이 이미 개발되어 있는데 이를 보다 발전적으로 활용할 수 있는 다양한 지원체계는 마련되어 있지 않다. 지원의 한 방법으로 연구데이터 전문가의 양성을 들 수 있다. 데이터 전문가는 기초학문자료센터 사업에서 DB를 구축하는 사업단의 안정적인 운영을

위해서 필수적이다. 연구데이터 전문가의 양성은 연구비 지원기관, 연구기관 및 관련 학문 분야가 협력하여 교육체계 및 과정의 공동 개발을 통해 자관의 시스템에 가장 부합하는 전문 인력을 양성해 낼 수 있을 것이다.

이와 같은 맥락에서 볼 때 한국연구재단 기초학문자료센터 인문학 분야 데이터의 수집 및 관리, 제공 방식은 인문학자들의 연구수행을 지원하는 데에 있어 매우 유용한 역할을 한다. 인문학 분야의 특수성을 고려한 연구데이터의 수집 및 관리가 이루어질 수 있도록 하는 것에 보다 초점을 맞추고 있으며, 인문학 분야에서 높은 이용빈도를 보이는 단행본, 학술지 등 연구에 기반이 된 핵심 문헌자료가 어떻게 사용되었는가에 대한 정보(맥락 정보)를 제공함에 따라 독자적으로 여러 문헌을 비교하고 해석하는 작업을 반복하는 인문학자의 물리적인 시간과 노력을 감소시키는데 상당한 도움을 줌으로써, 궁극적으로는 데이터의 공유 및 재이용을 원활하게 하여 활용성 증진에 기여한다고 할 수 있다.

### 5.3 연구데이터 활용체계 구축

기존의 연구자 성과평가는 연구결과물(특히 학술논문과 저서 등)의 생산 및 활용(대표적으로 인용)에만 초점이 맞춰져 있다. 향후 다양한 연구데이터의 관리, 공유 및 활용이 이루어지려면 이들도 다양한 방법으로 성과평가에 포함될 필요가 있다. 연구데이터가 학술연구의 결과물로 인정받고 보다 체계적으로 공유되려면 첫째, 공신력 있는 데이터 리포지토리에 기탁되어야 하고, 둘째, 이렇게 기탁된 데이터를 일관성 있고 안정적으로 인용할 수 있는 시스템이 필요하

다. 전자에 대해서는 이미 다양한 분야별, 기관별 리포지토리가 포진되어 있고 그 수도 계속 증가하는 추세이다. 기초학문자료센터 또한 개별 분야의 리포지토리라기 보다는 범인문사회 분야 리포지토리로서 국내 연구에서 생산된 다양한 연구데이터 및 관련 자료가 체계적으로 구축된 시스템이다. 후자에 대해서는 분산된 리포지토리에 기탁된 데이터를 보다 효율적으로 찾을 수 있도록 전세계적인 연구데이터 리포지토리의 등록소 역할을 하는 re3data.org가 현재 구축되고 있다.

데이터의 인용은 연구데이터가 연구수행의 결과물로 인정받고 더 활발하게 공유 및 재이용 되는데 기여한다는 점에서 매우 중요하다. 과학 데이터에 식별자가 처음 부여되어 인용될 수 있게 된 것은 2004년 독일의 연구비 지원기관인 German Research Foundation(DFG)이 독일 국립과학기술도서관(German National Library of Science and Technology, TIB)을 통해 DOI 번호를 부여한 것에서 시작한다(Brase, Sens and Lautenschlager 2015). 현재 DataCite라는 국제적인 협력기구가 가장 영향력 있는 서비스로 자리를 잡고 있다. re3data.org 서비스도 DataCite의 일부로서 제공될 예정이다. DataCite 이외에도 독일의 지구 및 환경과학의 데이터 리포지토리인 PANGAEA, 사회과학 분야의 대표적인 데이터 리포지토리인 ICPSR (Inter-university Consortium for Political and Social Research) 등에서도 독자적인 인용 체계를 가지고 있다.

따라서 기초학문자료센터를 포함한 국내 연구데이터 리포지토리도 향후 학술적 가치가 있는 연구데이터에 대해서는 국제적인 표준식별

자를 부여함으로써 데이터가 보다 공식적으로 활용될 수 있는 체계를 구축할 필요가 있다. 같은 맥락에서 현재 학술지 및 학술논문에 대해 적용되는 전통적인 영향력지수, 인용 등을 보완하고 대체하는 대안적인 지표(altmetrics)가 논의되고 있다. 연구데이터에 대해서도 인용 이외에 생산 및 활용에 대한 성과가 인정될 수 있는 환경이 장기적으로 마련되어야 할 것이다.

학술연구 분야에서 데이터가 기존의 학술논문, 저서, 학술대회발표논문, 특허 등과 같은 수준에서 결과물로 인정받고 학술 커뮤니케이션 시스템에 안착되기까지는 아직 이르다. 하지만 위에 언급한 일련의 활용체계의 구축은 연구데이터가 학술 커뮤니케이션에서 본연의 자리를 찾아갈 수 있도록 하는데 중요한 기여를 할 것으로 기대된다.

## 6. 결 론

본 논문은 기초학문자료센터 사업 초기부터 인문학 분야를 담당하여 자료를 수집, 정리하여 데이터베이스를 구축하는 업무를 수행해 온 저자들이 내부적인 관점에서 인문학 분야의 특수성을 고려하여 실제 연구데이터의 수집, 활용을 위해 사용한 방법론을 소개하고 이를 통한 향후 개선책을 모색해보자 하였다. 이를 위해 기초학문자료센터에 대한 개관과 더불어 인문학 분야의 특성을 고려한 업무 프로세스를 과제분석과 원자료 정보 구축의 관점에서 설명하였다. 특히 원자료 정보 구축에 있어 디지털 지식지도의 구현을 통한 연구데이터의 공유와 재이용을 설명하고 있는 부분은 많은 시사점을 가진다.

물론 아직까지 디지털 지식지도의 구현이 초기 단계에 있어 많은 연구를 통해 확고한 체계를 갖추어야 하지만 향후 인문학 분야의 데이터 아카이브 혹은 데이터 큐레이션의 발전에 있어 연구 전 과정의 맥락적 이해와 자료간 연계를 표현하기 위한 개념적 도구로써 충분한 의의를 가진다 하겠다. 또한 기초학문자료센터의 목표와 운영 및 관리체계, 그리고 인문학 분야 DB 구축의 실제 사례를 살펴봄으로써 향후 인문학 분야의 원자료의 원활한 수집 및 활용성 제고를 위한 정책, 인프라 및 활용체계의 구축에 대한 대안 또한 제시하였다.

본고에서 제안하고 있는 연구데이터 수집 및 활용성 증진 방안은 다음과 같다.

첫째, 연구데이터의 수집과 공개 관련 정책의 시행이다. 한국연구재단의 경우 연구비 지원을 받아 수행된 연구에서 생산되는 데이터와 결과물을 재단에 기탁해야 한다는 규정을 이미 시행 중이기는 하지만 실제로 연구가 완료된 후에 연구성과물이 아닌 연구에서 수집, 생산된 데이터가 존재 및 기탁 여부를 확인하고 이를 지키지 않은 것에 대한 처리 등의 후속 관리에 대한 체계가 마련되어 있지 않다. 따라서 미국과 영국 등 데이터 아카이브 선진국에서 실시하고 있는 데이터관리계획(DMP)을 기초학문자료센터의 설정과 분야에 특성에 맞게 재구성하여 체계적인 데이터관리계획을 정책으로 마련한다면 동분야 연구데이터의 수집과 공유에 보다 효과적인 결과를 야기시킬 것이다.

둘째, 데이터 관리, 수집 및 공유를 위한 인프라 및 환경 구축이다. 데이터관리계획과 같은 정책이 실제로 시행된다 하더라도 이것이 연구

데이터의 관리 및 공유의 모든 문제를 해결해 주지는 않을 것이다. 결국 개별 연구자의 관점에서 연구데이터의 체계적인 관리와 공유가 정책적인 의무 사항이 아닌 정상적인 연구수행의 일부로 인식되고 이에 따른 행동으로 나타나야 한다. 따라서 데이터관리계획의 적용은 정상적인 연구의 일부로서 데이터의 관리 및 공유라는 점에 대한 지속적인 인식 개선 노력이 정책 당국 및 학술 커뮤니티에 의해 선행되어져야 한다. 이와 더불어 연구자들이 데이터를 체계적으로 관리하고 공유할 수 있도록 전문적으로 지원하는 인력(데이터전문가)의 양성 및 활용이 이루어 질 수 있는 기반을 마련하여야 한다. 인문사회 분야를 지원하는 기초학문자료센터와 같은 시스템 환경이 이미 개발되어 있는데 이를 보다 발전적으로 활용할 수 있는 다양한 지원체계가 미흡한 상황에서 연구데이터 전문가의 양성은 DB를 구축하는 사업단의 안정적인 운영을 위해서도 필수적이다.

셋째, 연구데이터 활용체계 구축이다. 다양한 연구데이터의 관리, 공유 및 활용이 이루어지려면 이들도 다양한 방법으로 성과평가에 포함될 필요가 있다. 연구데이터가 학술연구의 결과물로 인정받고 보다 체계적으로 공유되려면 우선 공신력 있는 데이터 리포지토리에 기탁되어야 하며, 기탁된 데이터를 일관성 있고 안정적으로 인용할 수 있는 시스템이 필요하다. 기초학문자료센터를 포함한 국내 연구데이터 리포지토리도 향후 학술적 가치가 있는 연구데이터에 대해서는 국제적인 표준식별자를 부여함으로써 데이터가 보다 공식적으로 활용될 수 있는 체계를 구축할 필요가 있다.

## 참 고 문 헌

- [1] 고영만. 2010. 『의미기반의 분석적 정보서비스를 위한 기초학문자료센터 연구성과물 DB의 지식구조 분석』. 대전: 한국연구재단.
- [2] 김기덕, 이상훈. 2002. 『인문학관련 영상자료 실태조사 및 인문학 영상 아카이브 구축 방안』. 서울: 한국교육개발원.
- [3] 김선미, 이나니. 2005. 대학 학술정보 관리를 위한 오픈 액세스 기반 기관 레포지터리 운영방안 연구. 『정보관리연구』, 36(2): 45-71.
- [4] 김선태 외. 2010. 과학데이터 보존 및 활용모델에 관한 연구. 『한국비블리아학회지』, 21(4): 81-93.
- [5] 김은정, 남태우. 2012. 연구데이터 수집에 영향을 미치는 요인 분석. 『정보관리학회지』, 29(2): 27-44.
- [6] 김지현. 2010. 교수들의 셀프 아카이빙 저작물의 종류와 저장소에 관한 연구. 『한국문헌정보학회지』, 44(1): 53-74.
- [7] 김지현. 2012. 대학 내 연구자들의 연구데이터 관리에 관한 연구. 『한국도서관·정보학회지』, 43(3): 433-455.
- [8] 김지현. 2013. 국외 정부연구비지원기관의 연구데이터 관리정책 분석. 『한국문헌정보학회지』, 47(3): 251-274.
- [9] 김현희 외. 2005. 국가지식정보의 효율적인 유통체계 구축을 위한 대학 리포지토리의 운영 모형 개발: dCollection을 중심으로. 『정보관리학회지』, 22(3): 103-127.
- [10] 김현희, 정경희, 김용호. 2006. 대학 기관 리포지토리의 운영 현황 분석 및 개선 방안에 관한 연구. 『정보관리학회지』, 23(4): 17-39.
- [11] 김현희 외. 2006. 경제사회 분야 정부 출연 기관의 오픈 액세스 기반 지식 공유 활성화 연구. 『한국도서관·정보학회지』, 37(2): 145-171.
- [12] 배경재. 2011. 『토대기초연구지원사업 성과확산 강화 방안 연구』. 대전: 한국연구재단.
- [13] 신영란. 2012. 『인문사회 연구데이터 아카이브의 발전 방안에 관한 연구』. 석사학위논문, 이화여자대학교 정책과학대학원.
- [14] 신영란, 정연경. 2012. 국내 인문사회 연구데이터 아카이브의 개선방안에 관한 연구. 『한국기록관리학회지』, 12(3): 93-115.
- [15] 안영희, 박옥화. 2010. 디지털 큐레이션 정책을 위한 프레임워크 개발. 『한국도서관·정보학회지』, 41(1): 167-186.
- [16] 이길우. 2007. 『연구관리전문기관 성과관리 실태분석 및 개선방안에 연구』. 서울: 한국과학기술기획평가원.
- [17] 이상환, 심원식. 2009. 『과학데이터의 공유와 활용』. 대전: 한국과학기술정보연구원.

- [18] 이수상 외. 2007. 『기초학문자료센터 오픈 액세스시스템 구축 및 운영 방안 연구』. 서울: 한국학술진흥재단.
- [19] 임건면. 2007. 『기초학문자료센터 운영을 위한 제도 연구』. 서울: 한국학술진흥재단.
- [20] 유소영. 2015. Cyberscholarship 환경에서의 인문학 분야 연구데이터 사용 패턴 및 인식에 관한 연구. 『한국문헌정보학회 학술발표논집』, 35: 9-28.
- [21] 윤정옥. 2009a. 인문학자의 정보자원 이용에 관한 연구. 『정보관리학회지』, 26(1): 7-34.
- [22] 윤정옥. 2009b. 인문학자의 전자정보원 이용행태에 관한 연구. 『한국문헌정보학회지』, 43(2): 5-28.
- [23] 장금연. 2004. 학술정보유통을 위한 레포지터리 적용 방안 연구. 『한국문헌정보학회지』, 38(4): 291-310.
- [24] 장진규. 2007. 『기초학문자료센터 구축사업의 정책적, 경제적 효과분석』. 서울: 한국학술진흥재단.
- [25] 정영미, 이상기. 2010. 성공적인 리포지터리의 운영정책에 관한 연구. 『정보관리학회지』, 27(4): 131-152.
- [26] 한국학술진흥재단. 2006. 『기초학문자료센터 구축 사업 기본계획』. 서울: 한국학술진흥재단.
- [27] 한국연구재단. 2015. 『2014년도 학술자원 공동활용체계 구축사업(기초학문자료센터) 결과보고서』. 대전: 한국연구재단.
- [28] 황혜경, 김혜선, 최선희. 2004. 연구정보자원으로서의 기관 레포지터리 구축 현황과 과제. 『한국비블리아 발표논집』, 10: 77-102.
- [29] 황혜경, 이지연. 2009. 오픈액세스기반 기관리포지터리 성공에 미치는 요인 분석. 『정보관리학회지』, 26(4): 35-57.
- [30] American Council on Learned Societies. 2006. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: American Council on Learned Societies.
- [31] Beagrie, N., Beagrie, R. and Rowlands, I. 2009. "Research Data Preservation and Access: The Views of Researchers." *Ariadne*, 60.
- [32] Beagrie, N. and Greenstein, D. 1998. *A Strategic Policy for Creating and Preserving Digital Collections: a report to the Digital Archiving Working Group*. London: British Library Research and Innovation Centre.
- [33] Borgman, C. L. 2012. "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology*, 63(6): 1059-1078.
- [34] Buckingham, S. S. 2006. Sense-making on the pragmatic Web: A Hypermedia Discourse Perspective. *Proc. PragWeb '06: 1st International Conference on the Pragmatic Web, Stuttgart, September*, 2006: 21-23.

- [35] Burrows, T. 2011. "Sharing humanities data for e-research - conceptual and technical issues." *Sustainable data from digital research: Humanities perspectives on digital scholarship*, 177-192.
- [36] Diekema, A. R., Wesolek, A. and Walters, C. D. 2014. "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories." *The Journal of Academic Librarianship*, 40(3): 322-331.
- [37] Gutmann, M. et al. 2004. "The selection, appraisal, and retention of social science data." *Data Science Journal*, 3: 209-221.
- [38] Hedstrom, M. and Niu, J. 2008. Incentives for Data Producers to Create "Archive-Ready" Data: Implications for Archives and Records Management. *In Proceedings from the Society of American Archivists Research Forum*.
- [39] Jacoby, J. 2010. "Share and share alike? Data-sharing practices in different disciplinary domains." *Social Science Libraries: Interdisciplinary Collections, Services, Networks*, 144: 79-94.
- [40] Oppenheim, C., Probets, S. and Gadd, E. 2003. *Project RoMEO final report*. JISC.
- [41] Palmer, C. L. and Neumann, L. J. 2002. "The information work of interdisciplinary humanities scholars - Exploration and translation." *The Library Quarterly*, 85-117.
- [42] Perry, C. M. 2008. "Archiving of publicly funded research data: A survey of Canadian researchers." *Government Information Quarterly*, 25(1): 133-148.
- [43] Pike, W. and Gahegan, M. 2007. "Beyond ontologies: Toward situated representations of scientific knowledge." *International Journal of Human-Computer Studies*, 65(7): 674-688.
- [44] Pilat, D. and Fukasaku, Y. 2007. "OECD principles and guidelines for access to research data from public funding." *Data Science Journal*, 6: OD4-OD11.
- [45] Schäfer, A. et al. 2011. *Baseline Report on Drivers and Barriers in Data Sharing*. Opportunities for Data Exchange (ODE).
- [46] Simpson, P. 2005. *TARDis Project Final Report*. University of Southampton.
- [47] Stone, S. 1982. "Humanities scholars - information needs and uses." *Journal of documentation*, 38(4): 292-313.
- [48] Swan, A. and Brown, S. 2005. *Open access self-archiving: An author study*. Key Perspectives Limited.
- [49] Swan, A. et al. 2005. "Delivery, management and access model for E-prints and open access journals within further and higher education." *Learned Publishing*, 18(1): 25-40.
- [50] Tahir, M., Mahmood, K. and Shafique, F. 2010. "Use of electronic information resources and facilities by humanities scholars." *The Electronic Library*, 28(1): 122-136.
- [51] Vardigan, M. and Whiteman, C. 2007. "ICPSR meets OAIS: applying the OAIS reference

- model to the social science archive context.” *Archival Science*, 7(1): 73-87.
- [52] Velterop, J. J. 2005. *Open access publishing and scholarly societies: a guide*. Open Society Institute.
- [53] Watson-Boone, R. 1994. “The information needs and habits of humanities scholars.” *RQ*, 34(2): 203-215.

• 국문 참고자료의 영어 표기

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

- [1] Ko, Young Man. 2010. *Study on Knowledge Organizing System of KRM DB for Analytic Information Service based on Semantic Relation*. Daejeon: National Research Foundation of Korea.
- [2] Kim, Ki-Duk and Lee, Sang-Hoon. 2002. *A Research on the Actual Condition of Audio-Visual Materials and the Establishment of an Audio-Visual Archive Center in Relation to the Humanities*. Seoul: Korean Educational Development Institute.
- [3] Kim, Sun-Mi and Lee, Na-Nee. 2005. “A Study on Building an Open Access Based Institutional Repository in the University for Scholarly Communication.” *Journal of Information Management*, 36(2): 45-71.
- [4] Kim, Sun-Tae et al. 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.
- [5] Kim, Eun-Jeong and Nam, Tae-Woo. 2012. “Factor Analysis of Effects on Research Data Collection.” *Journal of the Korea Society for Information Management*, 29(2): 27-44.
- [6] Kim, Ji-Hyun. 2010. “A Study on Types of Content and Venues for Faculty Self-archiving.” *Journal of the Korean Library and Information Science Society*, 44(1): 53-74.
- [7] Kim, Ji-Hyun. 2012. “A Study on University Researchers’ Data Management Practices.” *Journal of Korean Library and Information Science Society*, 43(3): 433-455.
- [8] Kim, Ji-Hyun. 2013. “An Analysis of Data Management Policies of Governmental Funding Agencies in the U.S., the U.K., Canada and Australia.” *Journal of Korean Library and Information Science Society*, 47(3): 251-274.
- [9] Kim, Hyun-Hee et al. 2005. “Implementing the Management Model of the Institutional Repositories for the Efficient Distribution Infrastructure of National Knowledge Focusing on the dCollection System.” *Journal of the Korea Society for Information Management*, 22(3): 103-127.

- [10] Kim, Hyun-Hee, Joung, Kyoung-Hee and Kim, Yong-Ho. 2006. "A Study on the Improvement of University Institutional Repositories (dCollection) based on its Current State." *Journal of the Korea Society for Information Management*, 23(4): 17-39.
- [11] Kim, Hyun-Hee et al. 2006. "Promoting Resource Sharing among Government-Sponsored Economic and Social Research Institutes Focusing on IKIS (Institute Knowledge Inventory System)." *Journal of Korean Library and Information Science Society*, 37(2): 145-171.
- [12] Bae, Kyung-Jae. 2011. *A Study on Schemes for Putting KRM to Good Use for Disseminating the Outcomes of Infrastructural Basic Research*. Daejeon: National Research Foundation of Korea.
- [13] Shin, Young-Ran. 2012. *A study on the Development Plan for Humanities and Social Sciences Research Data Archive*. M.A. thesis, The Graduate School of Policy Science, Ewha Womans University.
- [14] Shin, Young-Ran and Chung, Yeon-Kyoung. 2012. "A Study on the Improvement Plans of the Humanities and Social Sciences Research Data Archives in Korea." *Journal of Records Management & Archives Society of Korea*, 12(3): 93-115.
- [15] Ahn, Young-Hee and Park, Ok-Wha. 2010. "Development of a Framework for Digital Curation Policy." *Journal of Korean Library and Information Science Society*, 41(1): 167-186.
- [16] Lee, Kil-Woo. 2007. *A Study of Analyzing R&D Performance Management System and Improving on How to Manage of Special Agency of National R&D Management*. Seoul: Korea Institute of Science Technology Evaluation and Planning.
- [17] Lee, Sang-Hwan and Shim, Wonsik. 2009. *The Sharing and Using Scientific Data*. Daejeon: Korea Institute of Science and Technology Information.
- [18] Lee, Soo-Sang et al. 2007. *The Strategic Research for the Open Access System of Korean Research Memory*. Seoul: The Korea Research Foundation.
- [19] Im, Geon-Myeon. 2007. *A Study on the Legislative System for Efficient Management of the Korean Research Memory (KRM)*. Seoul: The Korea Research Foundation.
- [20] Yu, So-Young. 2015. Cyberscholarship Hwangyeongeseoui Inmunhak Bunnya Yeongudeiteo Sayong Paeteon Mit Insige Gwanhan Yeon-gu. *Korean Society for Library and Information Science Occasional Papers Series*, 35: 9-28.
- [21] Yoon, Cheng-Ok. 2009a. "A Research on the Use of Information Resources of the Humanities Scholars in Korea." *Journal of the Korea Society for Information Management*, 26(1): 7-34.
- [22] Yoon, Cheng-Ok. 2009b. "Research on the Usage of Electronic Information Resources of the Humanities Scholars in Korea." *Journal of Korean Library and Information Science Society*,

- 43(2): 5-28.
- [23] Jang, Kum-Yeoun. 2004. "A Study on Repository Construction Scheme for Scholarly Communication." *Journal of Korean Library and Information Science Society*, 38(4): 291-310.
  - [24] Chang, Jin-Kyu. 2007. *Gichohangmunjaryosenteo Guchuksaeobui Jeongchaekjeok, Gyeongjejeok Hyogwabunseok*. Seoul: The Korea Research Foundation.
  - [25] Jung, Youngmi and Lee, Sang-Gi. 2010. "A Study on the Operational Policies and Best Practice of Repository." *Journal of the Korea Society for Information Management*, 27(4): 131-152.
  - [26] The Korea Research Foundation. 2006. *Gichohangmunjaryosenteo Guchuk Saeop Gibongyehoek*. Seoul: The Korea Research Foundation.
  - [27] National Research Foundation of Korea. 2015. *2014 Nyeondo Haksuljawon Gongdonghwallyongchegye Guchuksaeop (Gichohangmunjaryosenteo) Gyeolgwabogoseo*. Daejeon: National Research Foundation of Korea
  - [28] Hwang, Hye-Kyong, Kim, Hye-Sun and Choi, Seon-Heui. 2004. Yeongujeongbojawoneuroseou Gigwan Repojiteo-ri Guchuk Hyeonhwanggwaa Gwaje. *Korean Biblia Society for Library and Information Science Occasional Papers Series*, 10: 77-102.
  - [29] Hwang, Hye-Kyong and Lee, Jee-Yeon. 2009. "Analyzing the Factors Affecting the Successful Deployment of the Open Source Based Institutional Repositories." *Journal of the Korea Society for Information Management*, 26(4): 35-57.

