

오픈아카이브 기반의 원자료 서비스: 기초학문센터를 중심으로

Raw Data Services Based on Open Archives of KRM(Korea Research Memory)

이상철, 충남대학교 문헌정보학과, shout237@nrf.go.kr

이응봉, 충남대학교 문헌정보학과, eblee@cnu.ac.kr

Sang-Chul Lee, Graduate School of Library & Information Science.

Eung-Bong, Lee, Prof. Dept of LIS, Chungnam National University.

원자료는 행하여진 연구결과를 도출하기까지 발생된 기초자료로써 다듬어지지 않은 자료를 의미한다. 이러한 원자료는 이미 행하여진 연구와 유사한 연구를 진행하는 후속연구자에게 있어 시간과 노력을 절약하는 것이 가능하게 해준다. 본 연구는 인문사회분야의 원자료 제공 서비스인 한국연구재단의 기초학문자료센터에 대한 문헌연구와 해외의 유사기관에 대하여 알아보고 지속가능한 서비스를 위한 방법인 오픈아카이브에 대하여 알아보고자 하였다. 향후 기초학문자료센터의 기 구축된 자료에 대하여 실제 사용자의 만족도를 조사하고 이를 통해 콘텐츠의 보완 및 메타데이터 항목의 적절성과 질에 대한 실질적인 연구가 필요하다.

1. 서 론

인문사회분야 연구에 활용되는 데이터는 과학기술분야와는 달리 그 중간 산출물의 형태가 정형화 되어있지 않고 수치로 나타내기가 어렵다. 때문에 그간의 원자료는 개인 연구자가 소장하거나 연구가 종료됨과 함께 사장됨으로써 이용할 수 없게 되어 왔다.

기초학문자료센터는 인문사회분야의 중간산출물을 수집하고 FRBR 기반의 메타데이터를 구축하여 체계적으로 관리하고자 설립한 기관 리포지터리이다. 기초학문자료센터는 지금까지 과거에 종료된 연구과제 데이터들을 수집하고 구축하여 왔으나, 2010년부터 전년도(09년) 연구과제에 대한 DB구축을 시작했다. 즉, 센터에서 성격이 DB구축에서 서비스로 전환되는 시점이 도래한 것이다.

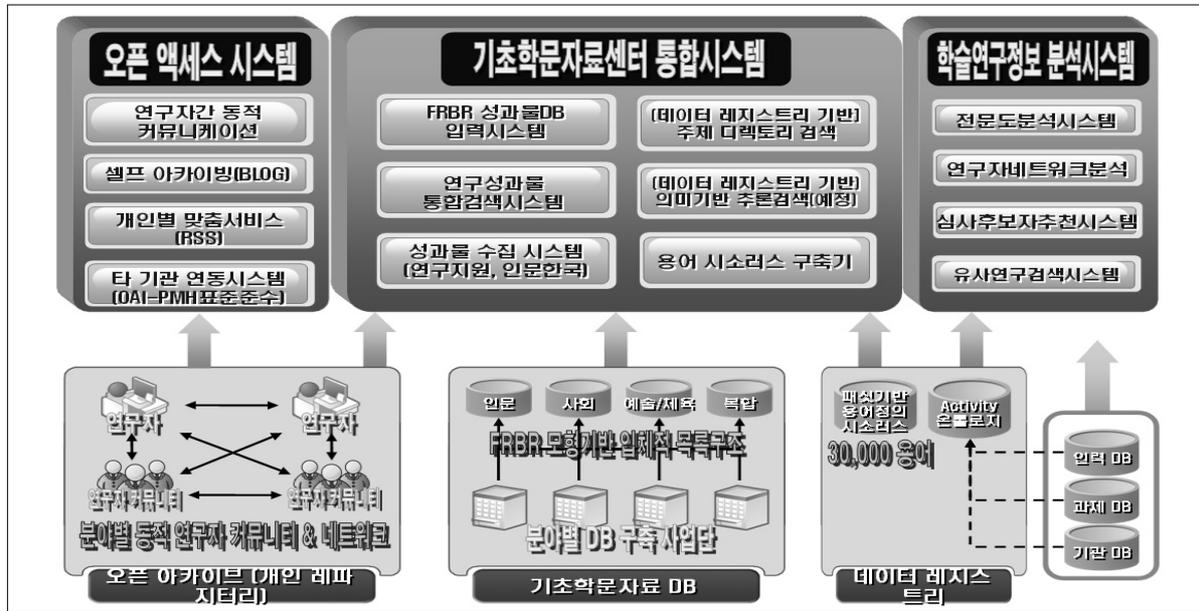
본 연구는 이미 해외의 많은 리포지터리가

추구하는 오픈아카이빙을 활용한 DB구축에 대하여 알아보고, 현재 기초학문자료센터에서 취하고 있는 오픈 아카이빙 방식에 대한 문제를 제기하며, 나아가 장기적인 서비스를 제공할 수 있는 모델을 제시하고자 한다.

2. 기초학문자료센터

구 학술진흥재단에서 출발한 기초학문자료센터는 2005년 시범사업을 시행한 이후 꾸준히 사업단을 운영해 오으로써 인문사회(예술체육 분야 포함)분야의 원자료를 구축하는데 힘써왔다. 현재 한국연구재단 운영 하에 2009년까지 약 60여만건 이상의 메타데이터를 구축하였고 이에 대한 디지털 원문자료까지 확보한 상태이다.

기초학문자료센터는 인문사회분야의 연구를 수행하며 얻을 수 있는 다양한 형태의 자료(보



<그림 1> 기초학문자료센터 서비스 모형도(한국연구재단 2009)

고서, 학술논문, 단행본, 기사날장, 고문서, 고도서, 동영상, 사진, 녹음자료, 통계 등의 조사자료, 웹문서 등)를 수집하고 이에 대한 메타데이터를 다수의 DB구축사업단을 운영하고 있다.

기초학문자료센터는 <그림 1>에서와 같이 인문사회의 각 분야별 DB구축사업단을 선정하고, 이 사업단에서 수집, 구축한 DB를 연구재단의 다른 시스템과 연계하여 서비스하는 역할을 하고 있다.

FRBR 개념을 도입한 메타데이터 항목은 이용자의 자유로운 항해와 다양한 접근점을 제공하는 것을 목표로 하고 있으며, 전문용어 데이터 레지스트리를 구축함으로써 정확한 정보에의 접근성도 확보하고자 하였다. 또한 과제 수행자의 연구업적을 통해 유사연구로의 항해도 가능하도록 하였다.

기초학문자료센터는 국가적 차원의 연구산출물 및 관련 기초자료 디지털 아카이빙 시스템 구축은 민간 부분이 할 수 없는 전형적인 시장실패의 영역으로, 매우 공공적인 성격의 사업이므로 정부가 공공사업으로 추진할 필요가 있는 바, 이의 구축을 통해 기초학문분야

전문연구자에 대한 학술적 가치가 높은 일차자료(원자료, 사실정보 등)의 제공을 통해 후속 연구의 활성화를 유도할 필요가 있는 것으로 판단하고 있다(김영기 2009)는 견해를 잘 반영한 것 이라 할 수 있다.

3. 국내·외의 자료센터

기초학문자료센터와 유사한 자료센터는 국내외의 다양한 사례로 찾아볼 수 있다.

국외의 사회과학분야의 리포지터리는 영국의 ESDS(Economic and Social Data Service), 독일의 GESIS(German Social Science Infrastructure Services), 미국 미시간대학의 ICPSR(The Inter-University Consortium for Political and Social Research) 등이 있고, 인문분야의 경우는 미국의 HRAF(The Human Relations Area Files), Project Gutenberg), 영국의 AHDS(The Arts and Humanities Data Service) 등이 있다.

국내의 경우는 국사편찬위원회의 한국역사통합정보시스템, 통계청의 국가 통계 포털 서

비스, 한국사회과학자료원, 한국사회과학데이터센터 등이 있다.

위의 리포지터리들의 주제분야 및 특성 등은 다음과 같다(표 1 참조).

<표 1> 국내·외 인문사회분야 자료센터

명칭	국가	주제분야	서비스
ESDS	英	사회과학	통계자료 질적조사자료
GESIS	獨	사회과학	국제비교조사자료 사회조사방법개발
IPCSR	美	사회과학	사회과학데이터 사회과학조사방법교육
HRAF	美	인류학 고고학	민속자료컬렉션 고고학자료
Project Gutenberg	美	어문학	연구자의 기고를 통해 저작권소멸 텍스트 디지털화
AHDS	英	고고학 인문학	영국국립자료 연표, 유물, 센서스, 경제지표, 미술·문학 작품, 공연 등 다수기관통합
한국 역사정보통합 시스템	韓	역사학	서지정보 편년자료 온라인커뮤니티
국가통계 포털서비스	韓	전분야	통계청제공의 각종 통계자료
한국 사회과학 자료원	韓	사회과학	사회과학 원자료 (양적·질적) 사회과학 문헌자료
한국 사회과학 데이터센터	韓	사회과학	사실정보 여론조사 출판 교육

특히 국외의 자료센터들의 경우에는 대부분 사회과학 조사방법론, 데이터 분석방법, 온라인 정보탐색기법 등 다양한 교육내용을 준비하여 이용자들에게 제공하고 있음을 알 수가 있다.

4. Open Archive

국가차원에서 지원하는 기초학문자료센터는 공공의 이익을 목표로 하는 바 오픈아카이브를 지향하고 있다.

이는 지속적인 서비스를 가능하게 하는 것으로서 구축과 이용의 선순환에서는 가장 바람직한 모형이라 할 수 있겠다.

우선 오픈아카이빙은 심사 완료된 연구논문을 OAI 프로토콜을 준수하는 리포지터리에 연구자 스스로 저장하는 것을 말한다. 이 경우 이용자는 정보를 탐색할 때 리포지터리의 위치와 검색위치 등을 신경 쓰지 않아도 되는 이점이 있다(이수상 2007).

기초학문자료센터는 한정된 예산 이상의 효과를 기대하기 위해 연구 참여자의 자발적인 오픈 아카이빙을 장려하고 있다. 한국연구재단의 과제를 수행한 연구자는 종료 후 결과물 및 중간단계 산출물을 스스로 기초학문자료센터에 등록하도록 유도하는 것이다.

실제로 2007년도부터 시작하는 과제에 대하여는 의무적으로 자료를 제공하도록 관련 규정을 개정하였으며, 저작권은 연구자 자신에게 있으나, 이를 서비스할 권리를 한국연구재단이 갖는다는 내용으로 구성이 되어있다.

이러한 시도는 몇 가지 측면에서 매우 바람직하다고 할 수 있다. ① 해당연구의 전문가인 연구자에 의해 제공된 정보는 전문성을 갖는다. 연구자는 해당연구를 수행하면서 자료가 어느 부분에 유용할 것인지를 잘 알고 있다. 때문에 보다 충실하게 메타데이터를 작성할 수 있으며, 메타데이터의 질적인 측면에 있어서도 전문연구자의 관점에서 서술할 수 있을 것이다. ② 연구자가 직접 데이터를 제공하도록 하는 것은 데이터의 중간 소실을 막는 역할을 할 수 있다. 현행 DB구축 방법으로는 연구수행이 끝난 후 기간이 어느 정도 지난 과제에 대하여 수집이 이루어지기 때문에

해당 기간 동안 자료의 손·망실에 대하여 대비할 수가 없다. 또한 어느 자료가 어느 시점에서 발생하였는지 정확한 파악이 어렵기 때문에 망라적인 자료수집에도 한계가 있다. 그러나 직접 연구에 참여한 연구자가 자료를 제출할 수 있도록 한다면, 나열한 문제점의 상당 부분을 해결할 수 있으리라 기대된다. ③ 연구자 스스로 자료를 제공하는 것은 비용 측면에서 효과가 있다. 현재 기초학문자료센터의 DB 구축은 몇 개의 DB구축사업단에게 맡기고 있는데, 여기에는 사업단 운영비 등의 비용이 발생하게 된다. '09년 기준 9개의 사업단을 운영하였고 이는 약 16억원의 예산을 사용하도록 하였다. 오픈아카이브는 사업단에 배정된 예산을 자료제공시스템 및 이용에 사용할 수 있게 함으로써 보다 양질의 서비스를 제공하는 것이 가능하도록 한다.

5. Open Archive 저해 요소

오픈아카이브는 여러 가지 측면에서 긍정적이지만 이것이 실제로 운영되는 데는 많은 어려움이 있다. 물론 한국연구재단의 지원과제의 경우에는 규정으로써 자료의 제출을 의무화하려 하지만 인문사회분야의 망라적인 자료를 구축하고자 하는 기초학문자료센터의 취지를 볼 때 아직까지 제약이 많이 따르는 것이 사실이다.

우선 참여, 공유, 개방이라는 Web 2.0의 정신이 우리나라에는 아직 뿌리 깊게 자리 잡지 못했다는 이유가 가장 크다. 해외의 경우, 전 세계 과학 및 의학 분야 문헌이 공공자원이 되어야 한다는 기본 취지에서 시작한 PLoS만 본다 하더라도 지식의 공유를 주장하는 움직임이 빈번하게 이루어지는 경우를 볼 수 있다. 그러나 우리사회의 실정은 자신이 이루어 낸 성과물을 공개하는 것을 꺼리는 것이 일반적인 현상이다. 또한 연구자 스스로는 공개하

는 것을 원할지라도 연구를 의뢰, 지원한 모기관의 입장에서 원자료 공개를 꺼리는 경우도 생각해 볼 수 있다.

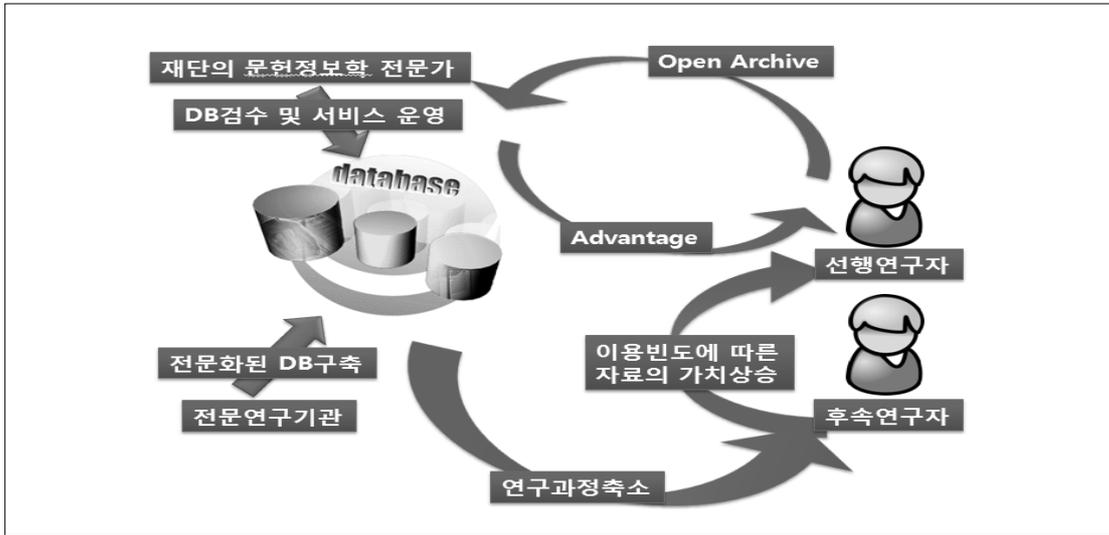
또 하나의 문제점은 원자료를 제공하는 연구자 입장에서 공개에 따른 충분한 보상을 받지 못한다고 느낄 수 있다. 이것은 아직 자료 제공에 의한 체계적인 보상정책이 수립되지 않은데서 기인하는 것으로 학자의 양심내지 막연한 후학도들에 대한 배려라는 명목으로 노력의 산출물을 제공하기 원하는 데는 무리가 있다.

마지막으로 정보관리에 대한 기술 부족을 들 수 있다. 해당 자료의 주제분야에 있어 연구자는 가장 적합한 전문가일 수는 있지만, 정보관리의 부분까지 전문성을 발휘해 주길 기대하는 것은 무리가 있다고 본다. 때문에 메타데이터의 충실한 내용에 비해 미흡한 데이터간의 연계와 조직구성으로 자칫 그 가치를 잃어버리게 될 수도 있다.

6. 장기적 운영을 위한 방안 모색

현재 존재하는 국내의 리포지터리는 그 실효성과 더불어 이용률 등의 문제로 인해 상당수가 규모의 축소내지 사장되는 위기에 처해 있다.

현재 구축된 기초학문자료센터의 데이터 중에는 이미 발간된 형태의 논문 및 보고서, 단행본이 상당수를 차지하고 있는데 이보다는 원래의 취지대로 원자료를 보존하고 이용하는 데이터센터로서의 역할에 중점을 두어야 할 것이다(석현호 2010). 이는 정체성의 문제뿐만 아니라 DB의 중복구축 측면에서 시간적, 물리적, 비용적 손실이 발생하게 된다. 따라서 기초학문자료센터는 원자료의 구축에 충실하되 후속연구자의 활용성을 고려하여 출간이 완료된 자료를 구축하는 것이 바람직하다고 본다.



<그림 2> 기초학문자료센터 운영모델

2010년에 진행된 관련 워크샵에서는 DB구축사업단의 운영방안에 대한 문제가 거론되었었다. 이는 현행 1년 단위 사업단 운영으로는 꾸준한 데이터의 질적 유지 및 메타데이터의 일관성을 유지하기 힘들다는 것이다. 따라서 인문사회분야에 전문성을 가진 기관과의 연계 혹은 양성을 통해 보다 체계적인 DB구축시스템을 구성하는 것이 대안을 모색하여야겠다.

또한 너무나도 방대하고 모호한 인문사회분야의 범주가 문제시 되는데 DB구축 대상범주를 명확히 하여 진정 해당 전문가에게 DB구축을 맡기고 여기에 대한 검수 및 운영 관리는 문헌정보학 전공자가 담당하는 모형이 바람직할 것이다.

여기에 오픈아카이빙을 장려하고 지원할 수 있는 현실적인 제도의 마련도 급선무라고 할 수 있다. 앞서 예를 든 PLoS와 같은 경우는 자체적으로 상당한 권위를 가진 저널을 출판하는 것이 연구자들에게 큰 매력으로 다가오는 것이 사실이다. 기초학문자료센터는 비록 직접 저널을 출간하지는 않지만 센터의 운영주체가 한국연구재단인 만큼 오픈아카이빙을

하는 연구자에게 상응할만한 보상체계를 구축하는 것은 가능할 것이다

기초학문자료센터의 장기적인 운영을 위한 모델은 (그림 2)와 같이 나타낼 수 있다.

7. 결 론

현재 한국연구재단에서 기초학문자료센터의 DB구축 사업단 운영의 규모는 점차 축소되고 있다. 이는 곧 능동적으로 수집되어지는 DB의 양이 적어짐을 의미하며, 이렇게 되었을 때 기 구축된 자료만으로 서비스를 지속하는 것은 케케묵은 백과사전을 10년, 20년 사용하는 것보다 못할 수 있다.

데이터의 갱신 없이 서비스의 활용을 기대할 수 없으며, 이를 몇 개의 사업단이 한정된 전문성을 가지고 계속적으로 데이터를 구축하는 것에도 한계가 있다고 본다.

데이터베이스를 구축할 때 가장 중요시되는 것은 메타데이터의 질인데, 현재 운영방식으로 넓은 인문사회분야의 모든 전문성을 감당할 수 있을지 의문이다. 따라서 자신이 행한 연구를 가장 잘 알고 있는 연구자 스스로

가 메타데이터를 작성하도록 하고 메타데이터에 관한 정보학분야의 전문가가 메타데이터를 검수하는 일이 운영에 있어 필요하다.

기초학문자료센터는 앞서도 소개한 바와 같이 후속 연구자의 반복을 최소화시켜 주는 부분에 있어 큰 의의가 있다. 이에 연구자는 학문의 발전을 진정으로 도모하는 마음으로 오픈아카이브를 접하여야 할 것이다. 하지만 기초학문자료센터와 같은 데이터 리포지터리의 운영은 무조건적이고 헌신적인 연구자의 참여만을 기다릴 것이 아니라 의미 있는 보상을 제공함으로써 연구자에게 보다 의욕적인 참여와 공유를 기대하여야 할 것이다.

이에 기초학문자료센터를 비롯한 많은 기관의 리포지터리에서는 오픈아카이브에 대한 보다 체계적인 연구가 필요하다

아울러 연구결과물과는 다르게 소실되기 쉬운 원자료에 대한 기초학문자료센터의 역할을 강화하고 기 구축된 자료의 활용에 대하여도 노력할 가치가 있다.

본 연구에서는 기초학문자료센터를 예로 삼았으나, 나아가 인문사회분야에 속한 많은 리포지터리를 대상으로 한 연구가 더 많이 진행되어야 할 것이다.

또한 기초학문자료센터는 주된 이용자와 기존의 DB구축을 담당했던 DB구축사업단, 기초학문자료센터를 관리해온 재단담당자와의 설문 및 면담을 통해 현재의 문제점을 정확히 진단할 수 있다면, 보다 구체적인 해결책을 제시할 수 있을 것으로 본다.

참고문헌

- 김영기. 2009. 『한술자원 공동 활용 기반구축사업 개선 방안 연구』. 서울: 한국도서관·정보학회지.
- 석현호. 2010. 『기초학문자료센터의 역할과 발전 방향』. 대전: 한국연구재단.
- 이수상. 2007. 『기초학문자료센터 오픈 액세스 시스템구축 및 운영방안연구』. 서울: 한국학술진흥재단.
- 이정연 외. 『기초학문자료 메타데이터 설계 분석 및 온톨로지 적용 방안 연구』. 서울: 한국문헌정보학회지.
- 임건면. 2007. 『기초학문자료센터 운영을 위한 제도연구 - 기초학문자료센터의 경쟁력 확보를 위한 중장기 발전계획』. 서울: 한국학술진흥재단.
- 장금연. 2004. 『학술정보유통을 위한 레포지터리 적용 방안 연구』. 서울: 한국문헌정보학회지.
- Charles W. Bailey, Jr. 2006. "What Is Open Access?"
<<http://www.digital-scholarship.com/cwb/OALibraries2.pdf>> [cited 2007.2.3].
- Peter Suber. "Open Access Overview: Focusing on open access to peer-reviewed research articles and their preprints."
<<http://www.earlham.edu/~peters/fos/overview.htm>> [cited 2007.2.3].