사회과학 데이터 아카이브 서비스

신영란_한국사회과학자료원

1. 서비스의 배경 및 필요성

한국사회과학도서관은 2006년 한국사회과학자료원(Korea Social Science Data Archive 이하 'KOSSDA'라 함)으로 기관명을 변경하고 사회과학 데이터 아카이브 서비스를 시작하였다. 이 서비스를 시작하게 된 가장 큰 이유 중의 하나는 사회과학의 경험적 연구에 기반이 되는 원자료(Raw Data 이하 '자료'라 함)[1]에 대한 수요가 크게 증대하였기 때문이다.

당시 국내에서는 자료의 수요가 증대하였음에도 불구하고 사회과학 자료를 수집해서 보급해 주는 전문적인 서비스 기관이 없었다. 이와 달리 사회과학이 발전한 구미 선진국에서는 일찍부터 양질의 사회과학 자료를 발굴하여 보존하고, 보존된 자료가 연구와 교육에 활용될 수 있도록 체계적으로 자료를 서비스하는 전문 데이터 아카이브들이 발달해 왔다.

매년 각급 기관에서는 막대한 예산을 들여 사회, 문화, 복지, 교육 등 사회과학 전 영역에 걸쳐서 정책 결정이나 연구의 목적으로 연구조사사업을 실시하고 있다. 그리고 이 과정에서는 방대한 양의 자료가 산출되고 있다. 그러나 이와 같이 산출되는 자료들이 기록 관리의 중요성에 대한 인식 부족과 기술적 또는 제도적인 여러 가지 문제들로 인하여 제대로 활용되지 못하고 소실되는 경우가 빈번하다. 이는 국가 예산의 낭비이며 사회적 손실일 수 있다. 따라서 많은 노력과 비용을 투자하여 산출한 자료들이 사장되지 않고 재활용되기 위해서는 데이터 아카이브와 같은 전문기관이 필요하다.

이러한 시대적 배경과 필요성에 의해 출범한 KOSSDA는 전문 데이터 아카이브의 역할을 수행하고자 학술적으로 활용 가치가 높은 자료를 수집해서 소장 문헌들과 함께 서비스하기로 하였다. 이로써 대표적인 사회과학 전문도서관이라 할 수 있는 한국사회과학도서관이 제공하던 문헌 중심의 서비스와 연구 자료 서비스를 결합한 새로운 시너지(Synergy) 형태의 사회과학 데이터 아카이브 서비스를 시작하게 된 것이다.

2. 시스템 개발

KOSSDA는 사회과학 데이터 아카이브 서비스를 구현하기 위하여 철저한 사전 준비를 거쳐 시스템 개발에 착수하기로 결정하였다. 단계별로 진행된 개발 과정은 다음과 같다.

첫 번째 단계에서는 성공적으로 사업을 추진하기 위해 개발의 목표와 방향을 분명하게 설정하였다. 일차적으로 시스템 개발의 궁극적인 목표를 사회과학 분야의 한국 관련 연구 자료와 문헌의 통합 DB 구축과 제공으로 설정하였다. 이와 함께 안정적인 자료 이관, DB 구축, 장기 보존, 상호호환성을 고려

1) 원자료(Raw Data)는 연구 결과를 도출하기 위해 사용되는 기초자료로 양적 자료(Quantitative Data)와 질적 자료(Qualitative Data)로 구분할 수 있다. 이러한 자료는 연구의 일차적 토대가 된다는 점에서 연구의 결과물이라 할 수 있는 논문, 보고서 등의 학술 문헌들과 구분된다.

한 표준화, 그리고 전문적인 사회과학 서비스를 구현하기 위한 서비스 시스템의 개발을 중점과제로 선정하였다. 특히 기술적인 측면에서는 단순히 신기술을 따르는 기술 중심의 접근 보다는 서비스 중심으로 특화된 시스템을 구축하기로 개발의 기본 방향을 정하였다.

두 번째 단계에서는 사회과학 데이터 아카이브 연구를 수행하였는데, 이는 전문적이고 표준화된 시스템을 개발하기 위한 목적으로 실시하였다. KOSSDA는 가장 먼저 해외의 대표적인 데이터 아카이브에서 제공하고 있는 각종 서비스에 대한 연구와 디지털 아카이빙(Digital Archiving)에 대한 연구를 수행하였다. 그 다음으로는 단위 시스템을 개발하기 위해서 자료관리, 메타데이터, 시소러스, 온라인분석도구에 관한 연구를 진행하였다. 그리고 마지막으로 이와 같이 심층적으로 수행한 연구 결과를 바탕으로 DDI(Data Documentation Initiative), OAIS 참조모형(Reference Model for an Open Archival Information System), HASSET(Humanities and Social Science Electronic Thesaurus)과 UNESCO 시소러스, NESSTAR(Networked Social Science Tools and Resources)를 개발 시스템에 적용하기로 채택하였다.

마지막 세 번째 단계에서는 앞서 설정한 시스템 개발의 목표와 방향, 그리고 다각적인 연구를 통해 채택한 연구 결과를 토대로 본격적인 시스템 개발에 착수하였다. 우선 시스템의 구성 요소와 세부 기능 요건을 정의하였고, 이를 기준으로 구축 모형을 설계하였다. 이에 따른 개발 내역으로는 기존의 문헌관리시스템(LAS), 이용자관리시스템, 이용 및 기탁 시스템을 보완한 새로운 서비스 시스템을 개발하였다. 또한 신규 시스템으로 DDI를 기반으로 하는 자료관리시스템(CMS), HASSET과 UNESCO 시소러스를 기초한 사회과학 시소러스 시스템, NESSATR를 적용한 온라인통계분석시스템을 개발하였다. 이밖에도 조사자료 변수관리시스템, 사회과학 연구영역 카테고리시스템, 이용문헌관리시스템을 개발하였다.

3. 서비스 내용

KOSSDA가 개발한 시스템은 국내는 물론 국외에서도 유래가 없는 도서관 시스템에 데이터 아카이브 시스템을 결합한 새로운 시너지 형태의 시스템이다. KOSSDA는 이 시스템을 통해서 도서관에서 제공하던 단편적인 문헌 서비스에서 연구 자료 서비스로 확장한 사회과학 데이터 아카이브 서비스를 제공할 수 있게 된 것이다.

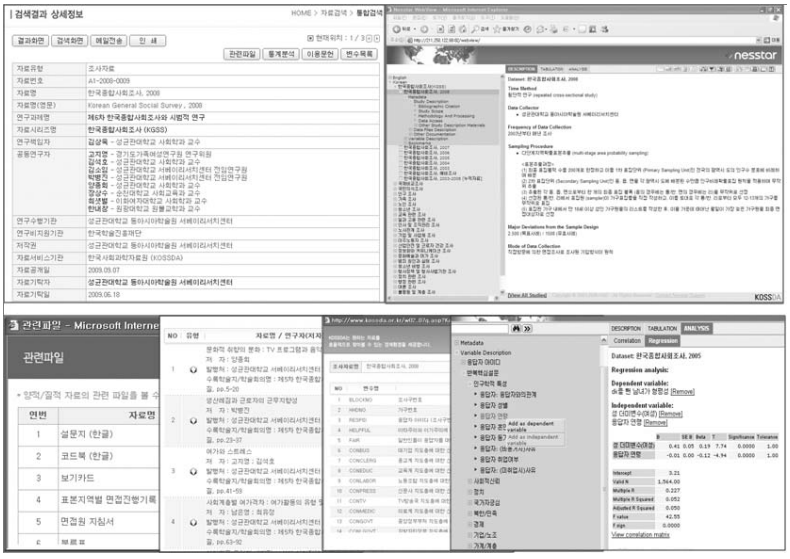
이 서비스의 기본적인 구성요소는 〈표 1〉과 같이 자료, 문서(Documentation), 그리고 출판물(Publication)이다. 자료는 조사자료 및 통계자료를 포함하는 양적 자료(Quantitative Data)와 기록문서, 관찰기록, 면접 및 구술 자료를 포함하는 질적 자료(Qualitative Data)로 구분하여 서비스하고 있다. 문서로는 조사자료의 설문지, 코드북 등과 질적 자료의 관찰기록노트, 녹취록 등을 포함하는 관련파일을 제공하고 있으며, 통계자료의 경우에는 통계표를 제공하고 있다. 출판물에는 연구의 결과물로 발간된 연구보고서를 비롯한 학술논문 등의 각종 문헌이 이에 해당하며 이용문헌이라는 유형으로 서비스하고 있다.

KOSSDA가 데이터 아카이브 서비스로 제공하고 있는 서비스의 종류는 크게 검색, 분석 및 탐색, 이용 및 기탁으로 구분할 수 있다. 대표적인 서비스 내용을 소개하자면, 검색과 관련된 서비

〈표 1〉 데이터 아카이브 서비스의 구성요소

구성요소	종 류		유형
자료	양적 자료	종합사회조사, 선거 · 여론 조사와 같은 표본조사, 센서스, 통계자료 등	조사자료 통계자료
	질적 자료	문화, 민속, 사회 등 분야의 각종 관찰, 면접, 구술 채록, 기록물 등	기록문서 관찰기록 면접/구술자료
문서	• 조사자료의 설문지, 코드북, 결과집계표, 코딩가이드 등 • 질적 자료의 관찰기록노트, 녹취록, 인터뷰 가이드 등		관련파일
	• 통계자료		통계표
출판물	• 연구의 결과물로 발간된 연구 보고서 • 학술지논문, 학위논문, 단행본 등		이용문헌

스로는 자료와 문헌의 통합검색, 연구분야별 검색, 시소러스 검색, 조사자료 변수검색을 제공하고 있다. 그리고 양적 및 질적 자료를 온라인상에서 분석해보거나 탐색해 볼 수 있도록 빈도분석을 비롯한 상관 및 회귀분석과 미리보기 기능을 적용하여 서비스하고 있다. 이용 및 기탁 서비스로는 검색과 분석 및 탐색을 통해 이용하고자 하는 자료를 신청하고 다운로드하는 기능과 문헌복사서비스 기능을 구현하여 제공하고 있다. 또한 온라인상에서 기관이나 개인이 소장하고 있는 자료를 기탁하고, 기탁한 자료에 대한 메타데이터를 간략 양식과 상세 양식 중에 선택하여 직접 입력할 수 있는 기능도 구현하였다(〈그림 1〉 참조).



〈그림 1〉 통계정보 알림 메일

지금까지 소개한 사회과학 데이터 아카이브 서비스의 특징은 먼저 DDI를 기반으로 상세하게 기술된 자료에 대한 정보를 검색 결과로 제공한다는 점이다. 또한 자료와 연관된 각종 관련 파일과 문헌을 동시에 검색할 수 있을 뿐만 아니라, 온라인상에서 분석과 탐색을 즉시 시행할 수 있다는 것도 이 서비스의

특징이다. 특히 통합 DB의 경우 사회과학 분야의 한국 관련 연구 자료와 문헌으로 그 대상과 범위를 특화해서 서비스한다는 점이 가장 두드러진 특징 중의 하나이다. 그리고 이를 사회과학의 주요 연구 분야별로 혹은 용어들 간의 관계로 확장해서 검색할 수 있도록 연구분야별 검색과 시소러스 검색을 구현하였는데, 이는 사회과학 주제 검색을 위한 도구로서 중요한 역할을 하고 있다. 이러한 여러 장점들 때문에 이 서비스는 연구자들로부터 유용한 서비스라는 긍정적인 평가를 받고 있다.

4. 서비스의 의의 및 향후 계획

KOSSDA는 사회과학 데이터 아카이브 서비스를 제공하기 위하여 국내에서는 유일하게 DDI 메타데이터를 적용하여 서비스를 구현하였다. 그리고 이 메타데이터 요소를 이용할 수 있는 온라인 통계분석 도구인 NESSTAR를 국내 최초로 도입하여 온라인상에서 양적 자료를 분석할 수 있게 하였다. 이러한 새로운 시도들이 바로 이 서비스의 의의라 할 수 있겠다. 그러나 무엇보다도 데이터 아카이브 서비스의 확산과 이를 통한 학문적_사회적 기여가 이 서비스의 가장 큰 의의라 할 수 있다.

아울러 자료관리시스템과 문헌관리시스템을 연동하여 연구자료와 문헌의 연계가 가능하도록 시스템을 개발함으로써 KOSSDA만의 특화된 서비스를 구현할 수 있었다는 점에도 중요한 의미를 부여할 수 있다. 최근 해외의 데이터 아카이브 서비스를 다시 조사해 본 결과, 단순히 자료만 서비스하던 아카이브들이 자료와 연관된 문헌을 함께 서비스하는 KOSSDA와 유사한 형태의 시스템으로 전환하고 있는 것을 확인할 수 있었다. 이는 결국 자료와 문헌은 분리해서 별개의 서비스로 구현하여 독립적으로 제공하기보다는 통합하여 함께 서비스할 때 좀 더 강력한 시너지 효과를 갖는 서비스가 될 수 있음을 시사한다.

최근 디지털도서관의 동향을 보면, 공유, 참여, 개방의 플랫폼을 기반으로 하는 웹 2.0 기술을 적용한 도서관 서비스 사례가 증가하고 있다. 이 뿐만 아니라 웹 2.0을 넘어서 웹 3.0의 기술을 도입한 소셜 시맨틱 디지털도서관(Social Semantic Digital Library, SSDL), 모바일 도서관(Mobile Library), 링키드 도서관(Linked Library) 등으로 표현되는 라이브러리 3.0 서비스 사례도 발표되고 있다. 개방과 공유를 강조하는 KOSSDA도 이미 이러한 기술들을 일부 서비스에 적용한 상태이다. KOSSDA는 앞으로 웹 2.0을 비롯한 웹 3.0의 기술을 점차적으로 확대하여 데이터 아카이브 서비스에 적용해 나갈 계획이다. 이를 통해 보다 상위 개념의 이용자 지향적인 플랫폼으로 이 서비스를 개선하고 다양한 목적으로 광범위하게 활용할 수 있도록 지속적으로 발전시켜 나가고자 한다.