오스트레일리아의 과학데이터 서비스체제(ANDS) 분석과 시사점

박동진†

요 약

현재 국가적 수준에서 과학 데이터세트에 대한 관리체제나 보전을 위한 구체적인 방침이 없다. 그래서 연구 프로젝트를 실시하고 있는 과학자 및 연구그룹들은 데이터세트에 대한 정보의 검색은 물론 공유가 불가능하다. 디지털화된 데이터를 이용하는 연구가 급격히 증가되고 있는 현 상황에서 연구자들 간에 과학데이터를 공유하고 재사용하는 것은 매우 중요하게 인식된다. 따라서 국가수준의 과학데이터 정책수립의 필요성이 대두되고 있다. 본 연구는 외국의 선진사례를 분석함으로써 우리나라의 전략적 계획수립에 있어서 중요한 시사점을 찾는 것이다. 먼저 과학데이터에 대한 일반적 사항과 국가별로 개략적인 과학데이터 정책방향을 살펴본 후, 우리나라와 비슷한 정부주도의 집중화된 연구 환경, 연구 지원체제 및 정보서비스 등으로 구성된 오스트레일리아를 대상으로 집중적으로 연구하였다. 구체적으로 ANDS(Australian National Data Service)를 분석하고 우리나라에 적용할 수 있는 시사점들을 도출하였다. 마지막으로 우리나라의 과학데이터 정책수립 시 반영되어야 할 가장 기본적인 원칙 들을 제시한다.

주제어 : 오스트레일리아 ANDS, 과학데이터, 전략계획, 국가수준, 원칙

Analysis and Implications of Australian National Data Service(ANDS)

Dong-Jin Park†

ABSTRACT

Our country does not currently have a concrete policy for the management and preservation of the scientific dataset on the national level. The scientists and the research groups that are implementing a research project are not capable of searching or sharing the information about the dataset. In this situation where there is a major increase in the number of researches that use digitalized dataset, being able to share and reuse the scientific data amongst researchers is recognized to be very important. Therefore our country needs a new formulated policy that manages scientific data on the national level. This paper helps to find the implications of the strategic planning in our country by analyzing previous advanced case studies done by foreign countries. We selected Australia as our subject because its intensive government-driven research environment, research infrastructure and information service are very similar to Korea. To be specific, we analyzed ANDS (Australian National Data Service) and drew out the implications that could be applied to our country also. And finally we propose the basic principles that needs to be mirrored when formulating a policy on our country's scientific data.

Key words : Australian, ANDS, Scientific Dataset, National Level, Strategic Planning, Principles

† 공주대학교 산업시스템공학과 교수
논문접수: 2011년 4월 8일, 1차 수정을 거쳐, 심사완료: 2011년 6월 2일

1. 서 론

현대의 과학 기술 분야의 연구는 전형적으로 데이터 위주의 연구(data-intensive research)이다[3][5]. 컴퓨팅 파워의 획기적인 증대는 방대한 데이터를 생산하여 새로운 연구의 기회를 제공하거나 지금까지 풀지 못했던 여러 문제들을 해결하기도 한다. 그래서 연구자들은 연구의 많은 자원을 데이터를 생성하거나 확보하는데 투입하고 있다. 연구조사로부터 산출된 각종 측정데이터, 관찰데이터 그리고 시뮬레이션 데이터 등은 다른 연구성과물과 마찬가지로 중요한 연구 결과물로 간주되어야 하며, 이렇게 많은 자원이 투입되어 만들어진 각종 데이터세트는 연구기관 및 국가적으로도 중요한 자산인 것이다[2][9]. 그래서 연구 자산인 데이터에 대한 관리활동은 연구자 및 연구기관이 책임져야 하는 중요한 연구 활동 중의 하나가 된다.

그러나 현재 우리나라에서는 이러한 연구 데이터 세트에 대한 관리체제나 보전을 위한 구체적인 방침이 없어서 연구 프로젝트를 실시한 기관은 물론 외부에서도 데이터세트에 대한 정보의 검색과 공유가 불가능한 경우가 많다. 국가수준의 정보서비스 측면에서도 보면 현재 과학기술 분야의 연구논문, 보고서, 특허정보, 산업동향 및 산업표준 등에 대한 정보는 KISTI의 NDSL 등에서 서비스가 되고 있으나, 아직 연구자를 위한 과학 데이터의 서비스나 관련 정보는 제공되지 않고 있다. 그래서 연구자들은 직접 데이터를 생성하거나, 해외에서 구매하거나, 연구자들 간의 비공식적인 절차나 방법으로 연구에 필요한 데이터를 확보하기도 한다. 최근 디지털화된 데이터를 이용하는 연구가 급격히 증가되고 있는 상황에서 연구자들 간에 데이터를 공유하고 재사용하는 것은 매우 중요하게 인식되고 있다. 따라서 이러한 과학데이터 서비스는 차세대 R&D 정보서비스 시스템에 있어서 주요한 과제가 된다.

과학 데이터의 장기보전(long-term preservation)과 공유는 국가적인 관심사이다. 이는 연구자 개인 및 연구기관 만의 역할로 가능한 부분이 아니다. 즉 연구자, 연구기관, 연구비지원기관, 학술단체, 데이터 저장 및 정보서비스기관 등 관련자(기관)들의 협력과 조정 등이 체계적으로 이루어져야 가능하기 때문이다[3]. 따라서 각국에서는 과학 데이터의 공공성, 대용량, 보안, 국내외적인 분산 및 협업연구 등을 고려하면서 매우 체계적이고 신중하게 접근하고 있다. 특히 미국, 영국을 포함한 EU, 캐나다, 오스트레일리아 등은 전문 위원회를 구성하여 집중적으로 검토하고 계획을 수립하고 있다.

최근 우리나라에서도 국가수준의 과학데이터 관리의 필요성이 대두되었고 있다. 본 연구는 외국의 선진사례를 분석함으로써 우리나라의 전략적 계획수립에 있어서 중요한 시사점을 찾는 것이다. 특히 우리나라와 비슷한 정부주도의 집중화된 연구 환경, 연구지원체제 및 정보서비스 인프라기구 등으로 구성된 오스트레일리아가 추진하고 있는 접근방법은 우리에게 시사하는 바가 매우 크다고 볼 수 있다. 제2장에서는 연구의 배경으로 과학데이터 관리에 대한 일반적 사항을 검토하고, 제3장에서 오스트레일리아의 ANDS 구축사례를 분석하고 시사점을 제시한다. 마지막으로 결론을 맺는다.

2. 연구의 배경

2.1 과학데이터와 관리 대상

데이터 위주의 연구 환경에서는 문서로 된 연구결과물의 출판만으로는 불충분한 것으로 간주된다. 점차 독자는 데이터에 대한 접근과 아울러 관련된 계산모형 등에 대한 접근을 필요로 한다. 이를 위해서 데이터는 연구성과물의 일부분으로 장기적으로 관리되어야 할 필요가 있다. 이러한 주장은 여러 보고서를 통해서 알 수 있다[2][8]. 이상의 보고서에서는 연구데이터 결과물의 장기 관리 및 서비스를 위한 체계적인 솔루션의 필요성을 주장하고 있다.

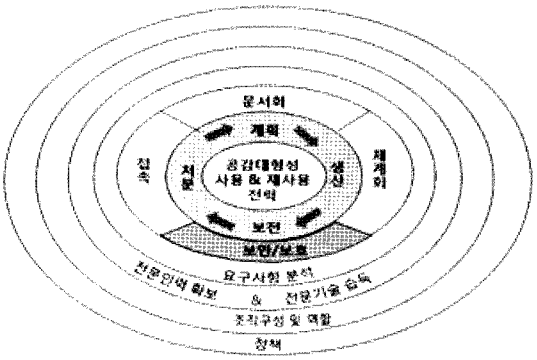
미국의 NSB(National Science Board)에 따르면 장기적 관리대상인 연구데이터(과학데이터와 혼용하여 사용함)들을 3종류로 구분한다[8]. 첫째, 관찰데이터(observation data)이다. 예를 들면 특정일자의 해양온도 측정값, 초신성의 폭발 등과 같이 다시는 확보할 수 없는 자료 등이 될 수 있다. 둘째, 실험데이터(experimental data)이다. 유전자 표현 패턴, 화학반응율 등이 가능한데 원칙적으로 조건이 정확히 알려져 있으면 데이터를 재생산 가능하나, 비용이나 알 수

없는 요인들이 영향을 미치는 경우에는 재생산이 불가능한 데이터이다. 마지막으로 계산 데이터(computational data)이다. 이것은 컴퓨터 모델이나 시뮬레이션 수행을 통해서 확보된 데이터이다. 모델에 관한 전체정보가 있으면 데이터의 재생산이 가능한 것이다.

이러한 연구데이터 모음이 관리의 대상으로 연구데이터 컬렉션(data collection) 혹은 데이터세트로 불리 운다. 이는 특정 집단에 의해서 모아진 연구데이터의 집합을 말하는 것으로 명확하게는 구분할 수 없으나 크게 3개의 유형으로 구분된다[8]. 첫째는 연구데이터 컬렉션(research data collections) 유형이다. 이것은 하나 혹은 그 이상의 연구프로젝트에서 산출된 연구데이터로서 사이즈에 있어서 차이는 있으나 상대적으로 작은 데이터의 모음으로, 데이터를 표현함에 있어서 표준양식은 없으며, 보관과 관리에 있어서도 체계적이지 않다. 또한 프로젝트의 종료 이후에 산출된 데이터 컬렉션에 대한 특별한 관리가 없는 것이 일반적이다. 둘째는 학문 및 연구분야 데이터 컬렉션(resource community data collections)이다. 이것은 특정 과학 혹은 공학집단에서 필요로 하는 데이터의 모음으로 커뮤니티에서 인정하는 표준체계에 의해서 데이터가 표현되고 관리된다. 데이터의 사이즈에 있어서는 연구데이터 컬렉션과 다음에 설명될 참조 데이터 컬렉션의 중간 규모이다. 이 데이터 컬렉션은 일반적으로 특정 학술 및 연구기관의 지원으로 운영되는데 장기적인 운영에 있어서 보장이 없는 경우가 많다. 셋째는 참조 데이터 컬렉션(reference data collections)이다. 이것은 과학 및 교육분야에서 많이 사용되는 참조 데이터로서 적용의 범위가 넓으며 다양한 학문, 기관, 국가 등에서 과학자, 학생, 교육자 등 다양한 분야의 사용자 그룹이 있다. 이 데이터는 표현의 기준이 엄격하고, 표준을 따라야 하며, 데이터 유지를 위한 예산이 확보되어 있으며 장기적으로 보관되고 관리된다.

2.2 과학데이터 라이프사이클 및 관리

미국의 과학기술자문회(NSTC: National Science and Technology Council)에 소속된 워킹그룹 중 하나인 IWGDD 에서는 [그림 1]과 같은 디지털 데이



[그림 1] 과학데이터 라이프 사이클과 관리모형(IWGDD Digital Data Life Cycle Model)

터의 라이프사이클 모델을 제시하였다[9]. 본 모델은 다소 포괄적인 의미지만 과학데이터의 라이프 사이클과 이와 관련한 관리모델을 제시해 준다는데 의미가 있다.

[그림 1]에서 보면 과학 기술 데이터는 계획(plan)-데이터생산(create)-보전(keep)-처분(disposition)의 라이프 사이클을 가진다. 이에 대하여 문서화(document), 체계화(organize), 보호(protect), 및 접속(access)의 네 가지 관리활동 직접 실시한다. 이러한 관리활동들의 외곽 원들에 해당하는 기술적 요구사항(technical requirements), 인력자원 및 전문기술 확보(human resources and professional skills), 조직구성(organizations and entities) 및 방침(policy) 등은 거시적인 관리활동으로 간주된다.

가장 핵심에 있는 과학데이터 라이프사이클을 보면, 계획단계에서는 연구목적으로 어떤 데이터를 생산해야 하는지 혹은 확보해야 하는지를 결정하는 것으로 이때 데이터의 라이프사이클 동안에 관리를 위해서 어떻게 행동을 취해야 하며 책임은 무엇인가를 파악하여야 한다. 생산/창조단계는 연구의 목적에 맞도록 데이터를 생산하거나 획득하는 것이다. 연구의 수행 중에 있으므로 연구목적에 충분하게 데이터를 보관하고, 관리하고, 접속하기 위해서 데이터를 저장한다. 이때 데이터를 활용하여 논문을 작성하거나 보고서를 제출한다. 보관단계는 데이터를 조직화하고 장기적으로 보관하는 것이다. 이때 기존의 데이터 컬렉션을 갱신하거나 추가하면서 데이터가 가능한한 오랫동안 유지될 수 있도록 주의를 기울인다. 과학데이터 라이프 사이클의 마지막 단계인 처분은 데이터를

이용한 연구의 목적을 달성하거나 더 이상 가치가 없다면 데이터를 폐기처분 하던지 혹은 또 다른 곳에서 활용할 가능성을 위하여 다른 조직에 이전하는 것이 말한다.

이상과 같은 과학데이터의 라이프 사이클 동안 크게 네 가지의 관리활동들이 라이프사이클 각 단계를 중첩해서 수행된다. 첫째, 문서화는 데이터 컨텐츠의 표준, 폼, 메타데이터, 갱신 횟수 등을 정의하고, 메타데이터를 만들어 유지하며, 데이터의 히스토리 관리 및 데이터를 처리하는 방법을 결정하는 것이다. 다음으로 체계화는 데이터 아키텍쳐, 처리, 구조들을 설계 및 구현하는 것으로 데이터 표준을 준수하는 것이 중요하다. 보안 및 보호는 데이터 품질 관리를 실행하고, 접속을 통한 통제의 수행, 재산권, 사용자 인증, 레파지토리(repository)의 보증 및 보안을 실시하는 것이다. 마지막으로 접속은 데이터 소스로부터 데이터의 접속이 가능하도록 하는 것이며, 데이터의 내용, 질, 가용성 등에 대한 카타로그를 작성하고, 분산된 데이터에 대한 식별성 제공하며, 가용한 데이터에 대한 정보 제공과 아울러 및 확산을 도모한다.

2.3 국가별 과학데이터 정책

2.3.1 국가수준의 정책방향

OECD의 주요 국가들은 과학데이터에 대한 엑세스 제공 서비스의 필요성을 인정하고 국가과학정책부문에서 위원회를 구성하여 정책의 방향을 설정하고 인프라 구축을 위한 노력을 기울이고 있다[12]. 각국은 연구인프라 개발을 위한 투자우선순위를 설정해 놓은 발표 문건들이 있는데 이러한 정책들은 주로 국가적 연구관리기관(예를 들면, 미국의 NSF과 국립과학원)에서 각 기관 및 전문가들의 협조로 개발된 것이다. 연구영역, 연구관련기관, 연구방법 등 매우 다양하므로 구체적이지 못하나 국가 전체적으로 방향을 제시하고 하위 정책수립의 필요성을 제시한다는 측면에서는 중요하다.

2.3.2 각국의 추진전략

미국은 이미 오래전부터 학문분야별 데이터 센터를 중심으로 과학데이터를 관리하고 있었으나 2006년부터는 국가과학기술자문회(NSTC)에서 IWGDD (Interagency Working Group on Digital Data)를 구성해서 국가수준의 과학데이터 관리 전략을 개발하고 있다[12]. 이를 기반으로 NSF의 Office of Cyber infrastructure에서는 5개의 컨소시엄들로 구성된 데이터네트를 계획하여 장기적으로 투자하고 있다. 특히 NSF 지원으로 이루어진 연구에서 생성되거나 수집된 데이터는 다른 연구자들에게 제공되어져야 한다고 명시하고 있다.

현재 과학데이터와 관련하여 가장 강한 리더쉽을 보이는 영국은 연합 프로젝트인 UKRDS(United Kingdom Research Data Service)를 통하여 많은 계획을 수립하고 있다. DCC(Digital Curation Centre)를 설립하여 관련 도구와 정책 및 컨설팅을 실시하고 있으며 RCUK, JISC, HEFCE, RIN, UKDA 등과 과학데이터 협조체제 구축하고 있다[12]. 영국은 국가수준에서 과학데이터 관리의 특성은 1) 장기적으로 일관성 있게 추진, 2) 서비스의 구현은 점진적인 접근, 3) 자율적인 참여를 기반하고 있다.

캐나다는 Research Data Canada 프로젝트가 정부 주도로 진행 중에 있으며, 대학, 연구기관, 도서관, 연구비지원기관, 연구자들로 구성된 Research Data Strategy Working Group(대학/연구기관/도서관/연구비지원기관/연구자 등의 대표자) 을 중심으로 체계적으로 전략이 수립하고 있다[12].

중국도 2002년부터 과학기술부에서 SDSP (Scientific Data Sharing Program)을 계획하여 계속적으로 과학데이터 관리체제를 구축하고 있다[7]. 중국과학원을 중심으로 2010년 말까지 40개의 과학데이터 센터, 300개의 마스터 데이터베이스 및 포털을 구축하였다.

2.4 국가별 추진의 일반적인 경향

2.4.1 국가수준의 위원회를 통한 전략수립

각국에서는 과학 데이터의 공공성, 대용량, 보안, 국내외적인 분산 및 협업연구 등을 고려하면서 매우 체계적이고 신중하게 접근하고 있다. 특히 미국, 영국을 포함한 EU, 캐나다, 오스트레일리아 등은 전문 위원회를 구성하여 집중적으로 검토하고 계획하였으며, 미국의 과학기술자문회(NSTC: National Science and

Technology Council)의 워킹그룹 중 하나인 IWGDD과 영국의 조인트 프로젝트인 The UK Research Data Service (UKRDS)와 호주 ADNS TWG 등과 같은 위원회를 두어서 국가수준의 과학 데이터 관리 체제의 구축을 위해 각국에 적합한 타당성 조사 및 계획을 실행하고, 단계적으로 접근하고 있다[12].

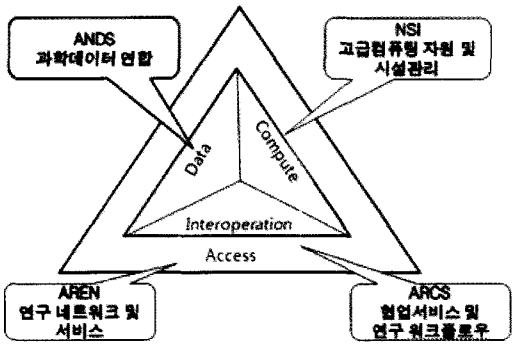
2.4.2 일반적 원칙의 적용

정책이 효과적이고 실현가능하기 위해서는 가능한 한 현실적인 연구 및 연구원들의 상황에 적합하여야 하므로 연구자금지원기관의 정책이나 연구기관의 정책은 해당국가의 연구지원 모델(research funding model)에 따라 결정되어야 하나 과학데이터의 일반적인 특성으로 인하여 모든 국가가 전제하는 보편적인 원칙이 존재한다. 즉 모든 국가에서 공통적으로 적용이 가능한 원칙이 존재하는 것을 알 수 있다.

3. ANDS 사례분석

3.1 추진전략

오스트레일리아의 국가연구기반구축사업인 NCRIS는 1997년부터 정부의 주도로 여러 협업 플랫폼 구축(Platforms for Collaboration) 프로그램들을 추진하였다. 2006년에 국가수준의 과학데이터관리와 관련한 컨설팅이 있었으며, 2007년에 대규모 포럼과 아울러 ANDS 기술워킹그룹(TWG: Technical Working Group)을 구성하고 연구를 통하여 보고서를 제출하였다[1][10][11]. 이를 기반으로 오스트레일리아의 DIISR와 EIF의 지원으로 Monash 대학을 중심으로 대학과 국가연구소가 협력하여 2008년 9월에 ANDS를 설립하고 운영하고 있다. ANDS는]그림 2]와 같은 국가적인 연구협력 체제하에서 과학데이터의 관리 및 공유를 위한 정책 및 서비스의 개발을 담당하고 있으며, 고급컴퓨팅 자원 및 시설을 관장하는 NSI, 협업서비스 및 워크플로우를 담당하는 ARCS 그리고 연구네트워크 서비스를 담당하는 AREN 등과 서로 공조한다[4]



[그림 2] 오스트레일리아의 국가협업연구 플랫폼

ANDS의 일관된 추진을 위해서 ANDS 조직위원회는 2008년도에 초기 비즈니스 계획(Interim Business Plan)을 수립하고 2009년에 계속적으로 수정하였다. 동시에 프로젝트 관리 위원회는 관련 단체들과 협약을 맺어 계속적인 지원과 협조를 획득하였다. ANDS에서는 오스트레일리아 국내에 흩어져 있으며 이질적이고 관리되고 있지 않는 과학데이터를 연구자원화 하기 위하여 향후 10년 동안 <표 1>과 같은 구체적 목표를 가진다(ANDS 2008).

<표 1> ANDS 목표

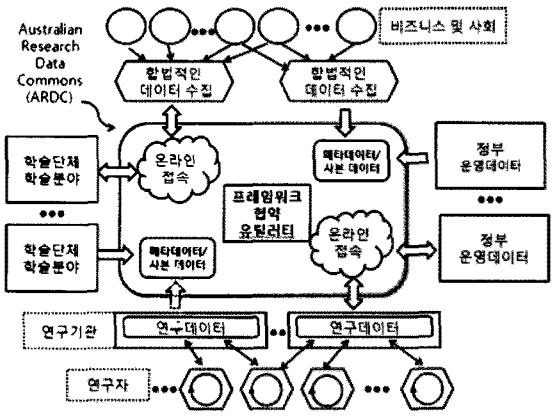
번호	대상	목표
1	연구 데이터 통합	국가적으로 모든 연구데이터가 데이터 공유체제(Data Commons)의 연구 레파지토리 네트워크에 포함되도록 한다.
2	연구 데이터 공유	국내의 연구자 및 연구데이터 관리자가 체계적인 데이터관리 방침 하에 세계 최고수준으로 데이터를 창조하고, 관리하고, 공유하도록 한다.
3	연구 데이터 저장	연구데이터가 안정된 체제와 환경하에서 주기적으로 저장, 접속, 관리되도록 한다.
4	전문성 확보	연구기관 및 관리기관의 많은 사람들이 데이터관리에 있어서 전문성을 갖도록 한다.
5	연구 데이터 검색	연구자들이 데이터 공유체제에 접속하여 적합한 데이터를 찾고 확보할 수 있도록 한다.
6	연구 데이터 사용	다른 분야 및 다른 연구자들의 데이터를 확보하고 교환하고 재사용하고 융합할 수 있도록 한다.
7	국제적인 데이터 교환	국가간에 데이터들을 서로 교류할 수 있도록 한다.

이상과 같은 역할을 하는 ANDS는 아래와 같은 7가지 원칙을 지키면서 추진되고 있다[1]. 첫째, 점진적 접근방법이다. ANDS는 데이터 저장, 기관들과의 연합, 서비스 등의 프레임워크를 구축하는데 있어서, 초기에는 소규모로 시작하고 점차 수요를 파악하면서 확장해 나가는 방법을 채택한다. 둘째, 정예 인력의 확보이다. ANDS는 연구데이터 통합의 비전에 공감하고 적극 동참하는 사람과 가시적인 결과를 보여줄 수 있는 사람들을 파악하고 이들과 먼저 작업을 실시한다. 셋째, 먼저 컨텐츠의 확보에 치중한다. 초창기에는 질적인 면에서 일정수준 이상의 데이터를 양적으로 충분히 확보하는 방법을 취하고 추후에 데이터의 질적인 문제에 초점을 맞추기로 한다. 넷째, 서비스 공급에 초점을 맞춘다. 초창기에는 연구와 데이터 분석 등의 업무보다는 서비스 제공을 위해서 필요한 기반의 구축업무를 우선적으로 고려한다. 다섯째, 전략적 파트너와 협력한다. 핵심사업인 연구데이터 공유체제(ARDC)를 구축하는데 있어서 협업연구를 지원하는 유관기관들(AAF, ARCS 등)과 협력한다. 여섯째, 실제 데이터의 보관 및 관리는 주로 연구데이터를 생산한 해당 연구기관에서 수행하도록 한다. 단지 ANDS는 각 기관의 저장소에 있는 데이터의 공개를 촉진하고, 제한적 접근 및 재사용이 가능하도록 돕는다. 또한 각 연구기관에 저장된 각 데이터들은 분야별로 표준화된 메타데이터 스키마로 기술하도록 한다. 일곱째, 연구데이터가 장기적으로 관리되도록 ANDS의 사업도 지속되도록 한다. ANDS의 업무계획은 3년 단위로 연속적으로 진행되도록 한다.

3.2 ANDS의 데이터 연합

전문가데이터 컬렉션은 하나의 물리적 공간에 있을 수도 있고 여러 개의 공간에 분산되어 있을 수도 있다. 또한 데이터 컬렉션 각각은 독특한 운영방침, 목적, 및 연구비 지원 기관의 요구 등에 따라 서로 다를 수 있다. 따라서 데이터 컬렉션 관리는 연구기관의 데이터 저장시스템의 범위 이상의 관점을 가진다. 그래서 ANDS에서는 연구데이터 관리를 위한 기반체제를 구축하는데 있어서 각 연구기관의 데이터 정보를 통합하는 데이터 연합(data federation)이라는 개념을 도입하였다[10]. 이는 물리적으로는 ANDS에서 연구데이터를 저장하고 있지는 않지만, 각 기관 및 연구원이 보유하고 있는 데이터에 대한 정보를 통합하여 관리함으로써 사용자가 데이터에 대한 정보를 파악하고 해당 데이터에 접속할 수 있도록 유도한다는 개념이다. 이러한 개념을 구체화한 실체가 연구데이터 공유체제(ARDC: Australian Research Data Commons)이다. 이것은 연구자가 데이터 컬렉션을 생산 및 저장하고, 네트워크로 접속하고 검색이 가능하도록 지원하는 정보기술 기능과, 데이터 전문가가 연구자들을 교육 및 컨설팅 하는 기능과, 데이터 공유와 확산을 지원하는 정책과 방침 등을 포함 하는 가상의 기반체제를 말한다.

[그림 3]은 ARDC의 개념도이다[10]. ARDC 정보의 소스는 첫째, 연구기관 그리고 연구자가 소유하고 있는 연구데이터이다. 둘째, 학술단체 그리고 혹은 학문분야에서 관리하는 연구데이터이다. 셋째, 비즈니스나 사회로부터의 합법적으로 수집한 연구데이터이다. 마지막으로 정부로부터 획득한 운영데이터이다. 실제 데이터는 원칙적으로는 외부기관이 보유하는 것이나 필요에 따라 데이터의 복사본을 ARDS 내에서 보관하기도 한다. ARDC에서는 단지 데이터에 대한 메타데이터를 레파지토리에 저장한다. 사용자는 온라인 접속을 통하여 ARDC 메타데이터에 접속하고 ARDC에서 제공하는 각종 서비스 유틸리티를 이용하여 연구기관에 저장된 실제 데이터를 확보한다. 이상의 모든 서비스가 가능하도록 ARDS 내에는 정보 및 연구데이터 서비스를 지원하기 위한 프레임워크, 협약, 유틸리티 서비스 등의 기능들이 포함된다. 이상의 기능들을 구현하기 위한 프로그램을 다음절에서 소개한다.



[그림 3] ARDC 개념도

3.3 ANDS의 실행 프로그램

이상에서 언급된 ARDC 비전과 개념도는 크게 4개의 상호 연관된 서비스 제공 프로그램으로 구분하고 각각을 위한 세부 프로젝트들을 계획하였다. 계획된 각 프로그램은 제한된 예산과 위험을 줄이기 위한 현실적인 접근으로 소수의 전문기관 및 집단에 의해서 수행되고 있다. 또한 각각의 프로젝트들은 연방정부와 국가연구기관과의 긴밀한 협조를 통해서 체계적으로 수행된다.

3.3.1 프레임워크 개발 프로그램

프레임워크 프로그램은 전반적인 방침을 설정하는 것으로, 연구자나 연구기관이 연구비 지원기관의 요구사항에 맞도록 어떻게 과학데이터를 보관하며, 이것을 ARDC에 어떻게 연계할 것인가를 정의하는 것이다. 데이터가 만들어지고 공유되면서 라이선스의 문제는 점차 단순화되고 줄어 들 것으로 예상한다. 프레임워크 프로그램의 주요 협력자들은 기관의 데이터 관리자, 정부기관의 담당자, 연구비제공기관, 연구기관에서의 분야별 리더와 연구지원부서 책임자들이 될 것이다. 프레임워크의 초점은 개별기관들 보다는 데이터관리 주체들 간의 상위수준의 연합(higher-level aggregation)이다.

3.3.2 유틸리티 제공 프로그램

이 프로그램은 데이터 서비스를 지원하는 프로그램이다. 예를 들면 학제적(cross-discipline) 검색서비스, 국가 데이터 컬렉션 등록서비스, ID 부여 서비스, 연합 등록서비스, 접속 방침 등록서비스 등이 된다. 사용자 접속(human-facing) 서비스는 AAF 인증 서비스와 ARCS 검증서비스를 채택한다. 언급한 각종 등록서비스는 데이터와 관련된 모든 정보들을 등록하는 것으로 매우 중요한 역할을 한다. 유틸리티 서비스를 개발하는 것과 아울러 ANDS에서는 국가수준의 연합 유틸리티 프레임워크(Federation Utilities National Framework)를 개발한다. 이것은 연구 학술단체 수준에서의 데이터 연합을 지원하는 것이다, 이것은 데이터 유틸리티 서비스의 호환성을 고려한 기술적인 섹터들과 공조하여 진행한다. 또한 프레임워크는 영국이나 미국과 공동으로 작업을 수행할 예정이다. 유틸리티 프로그램의 주요 수혜자는 ARDC의 콘텐츠 제공자와 소비자가 될 것이며, 데이터 시설관리자와 행정가 및 데이터 연합을 구축하고 운영하는 연구 커뮤니티들도 수혜자가 될 것이다.

3.3.3 ARDC 확산 프로그램

확산 프로그램은 ARDC에 더 많은 콘텐츠가 저장되도록 하고 ARDC를 더욱 활성화 시키는 구체적인 프로그램이다. ANDS를 위한 예산은 전체 연구섹터에 레파지토리 솔루션을 제공하기에는 충분하지 않다. 인건비는 매우 높으며 예산은 제한되었기 때문에 이 프로그램은 몇 개의 목표(target) 기관에 먼저 초점을 맞추어 지원하는 것을 지향한다. 이 프로그램은 목표기관 안에서 생산되는 데이터와 메타데이터가 파악되고, 저장되고, ARDS에 접속될 수 있도록 하는 것이다. ANDS는 데이터의 질적인 면과 양적인 것을 동시에 만족시키기 위하여 자동화 혹은 반자동화된 데이터와 메타데이터를 대상으로 삼았다. 특정 그룹을 선정하여 그들에게 예산을 지원해 주고 그들에게 장기간의 저장체계를 유지하도록 지원을 한다. 이와 관련하여 ANDS의 첫 해에는 스탭을 채용하고 훈련시키는데 집중하였으며, 가장 적절한 선택사항이 무엇인지 그리고 선택에 있어서 구체적인 기준을 개발하였다. 또한 첫 해에는 기존의 콘텐츠를 저장하는데 집중하고 있다. 기존의 유용한 컨텐츠를 파악하고 ARDC를 통하여 검색 및 정보제공이 가능하도록 할 것이다. 가능한 컨텐츠는 있지만 시스템이 없는 기관을 대상으로는, ANDS에서 소규모의 팀(small agile team)을 지원해 주어서 그들이 기존에 개발된 툴을 이용하여 데이터 저장체제를 갖추도록 지원해 줄 예정이다. 이미 저장체계가 되어있는 기관을 대상으로는 이 프로그램에서 이것들을 ARDC에 통합될 수 있도록 지원할 예정이다.

3.3.4 과학데이터 관리능력 배양 프로그램

이 프로그램은 국가적 차원에서 조직, 시스템, 서비스 그리고 인력 등에 대한 연구 데이터 관리 능력을 배양하는 것이 목적이다. 오스트레일리아의 e-research 와 정보기반구조(information infrastructure)

의 능력이 향상되도록 성숙도모델(capability maturity model)을 만드는데 것도 유관섹터와 공동으로 작업을 할 예정이다. 또한 ANDC에서는 관련기관과 함께 데이터관리와 관련하여 훈련 프로그램을 개발할 예정이며, 학술단체 등과 포럼 등을 통하여 폭넓게 관리 능력 배양 프로그램을 확산시킬 계획이다.

3.4 시사점

아직 어떠한 국가도 체계적으로 그리고 종합적으로 과학데이터관리 및 서비스를 완전하게 해결하고 있지 못하고 있으나 국가적인 중요성으로 인해 장기적인 시각으로 계속적으로 추진해 나가고 있음을 알 수 있다. 국가들 간에 상황과 추진하는 방법에는 다소 차이가 있으나 목표하는 방향은 거의 일치하기 때문에 과학데이터에 대한 기반연구가 부족한 우리나라의 입장에서는 벤치마킹이 그 무엇보다 중요하다. 특히 본 연구를 통하여 살펴본 오스트레일리아는 연구환경 및 정보인프라가 우리나라와 매우 흡사하고 현재 추진 중인 일련의 전략 및 방법들이 우리에게 아래와 같은 시사점을 준다.

첫째, 오스트레일리아의 경우에 ANDS가 발족하기까지 3년에 걸친 연구, 컨설팅, 검토 과정이 있었던 것은 주목할 만하다. 영국, 캐나다, 미국의 경우를 보더라도 과학데이터 서비스 제공은 기존의 학술정보서비스와 비슷하거나 그 이상의 준비기간을 가진 것으로 나타난다. 특히 각국에서는 관련단체의 폭넓은 전문가들을 중심으로 구성한 위원회에서 충분한 토의를 거쳐서 보고서를 제출하고, 이를 검토하고 승인한 후 예산을 배정하고 단계적으로 추진하였음을 알 수 있다.

둘째, ANDS의 핵심서비스인 ARDC는 데이터 연합(data federation) 개념의 가상의 공간임을 주목해야 한다. 즉 각 연구기관 및 연구자들이 실제로 데이터를 보유하고 있으며 ARDC는 이들의 정보를 보유하고 있는 기관과 연합하고 정보를 통합하여 서비스해 주는 것이다. 아직 초창기인 ANDS에서는 ARDC를 통한 서비스 자체보다는 데이터 연합을 위한 관련기관과의 협력, 제도의 완비, 표준의 설정, 법적인 문제의 해결 등 정보의 저장 및 통합 관점에 더 많은 비중을 두고 있다. 처음에는 시스템의 구현보다는 통합의 인프라구축에 더 많은 자원이 소요되어야 함을 알 수 있다.

셋째, ARDC에는 기본적으로 메타데이터 저장을 위한 레파지토리만 보유하고 실제 과학데이터 컬렉션은 주로 연구기관이나 학술단체 등에서 보유하는 방식을 채택한다. 따라서 실제 데이터를 생성하고 보유하는 연구기관의 협조와 이들을 위한 지원이 성공의 핵심이다. 앞으로 모든 연구기관에서는 외부 연구비를 지원받아 수행한 연구용역에 대한 근거로서 연구데이터의 보유가 의무화 될 것이므로, 이를 위한 준비 차원에서도 연구기관의 레파지토리의 구축방안에 대한 표준을 제시하고, 메타데이터 스키마의 표준 설정과 데이터 생성 및 연계 등에 대한 준비도 시급히 실행되어야 할 것이다.

넷째, ANDS는 10년의 장기목표를 설정하고 매 3년 주기로 예산을 재배정 하는 장기 프로그램이다. 즉 장기적 관점에서 원칙을 구체화해 놓고 이를 준수하며, 현실적이고 점진적 방법으로 접근한다. ANDS의 경우 4개의 실행 프로그램을 보면 현재의 수준에서 가능한 부분부터 하고 있는 것이다. 이는 영국을 비롯한 다른 국가에서 있어서도 마찬가지임을 알 수 있다. 특히 초창기에는 당장의 가시적인 결과에 연연하지 않고 인프라 구축에 더 많은 노력을 기울이고 있는 것 그리고 파트너의 선정도 현재 수준에서 충분히 준비된 기관이나 인력을 중심으로 단계적으로 접근하는 것을 주목할 필요가 있다.

마지막으로 ANDS의 최종 수혜는 결국 오스트레일리아 정부임을 알 수 있다. 즉 특정분야, 인접분야 혹은 융합분야 등에서 우수한 연구데이터를 공유하고, 재사용함으로써 연구의 생산성 향상과 아울러 비용의 절감을 가져오는 직접적인 효과가 있을 뿐 아니라, 연구의 수월성에 따른 양적/질적인 향상으로 인하여 궁극적으로 국가경쟁력을 제고하는데 기여한다는 것이다. 이러한 인식으로 인하여 선진 각국에서는 국가가 주도적으로 과학데이터를 관리하고자 각종의 사업을 추진하고 있는 것이다.

4. 결론 및 연구의 한계

디지털화된 데이터 중심의 연구가 급격히 증가되고 있는 상황에서 연구자들 간에 데이터를 공유하고

재사용 할 수 있게 하는 것은 국가차원에서 다루어져야 하는 중요한 과제로 인식되고 있다. 구미의 여러 국가들은 몇 년 전부터 체계적으로 준비하여 상당히 진전된 반면에 우리나라는 최근 들어 KISTI를 중심으로 과학데이터를 관리하고 서비스를 하기 위한 사업을 계획하고 있다. 이러한 때 이 분야의 선진국가에 대한 벤치마킹은 여러 면에서 중요한 의미가 있다. 본 연구에서는 먼저 과학데이터 관리에 관한 일반적인 사항을 정리하고 다음으로 이와 관련하여 오스트레일리아의 ANDS 추진내용을 살펴보았다.

본 연구에서는 결론으로 우리나라의 과학데이터 정책수립을 위하여 아래와 같은 세 가지를 제안한다. 첫째 과학데이터 정책수립을 위하여 국가차원의 여러 기관 및 단체의 적극적인 관여가 필요하다. 특히 정부 및 국가과학관련 위원회의 적극적인 협조 및 협력을 획득하고 안정적이며 장기적인 예산을 확보하며 단기적 성과에 연연하지 않는 것이 중요하다. 둘째, 정책이 장기적으로 일관성이 있게 추진되어야 하며 서비스 구현은 점진적인 접근을 취하는 것이 중요하다. 국가적으로 정책전담기관 및 실행전담기관들을 설치하고 정책개발, 인프라 구축 및 기술개발을 동시에 진행하며 단계적으로 통합하는 것이 바람직하다. 셋째, 계획수립은 하향식(top-down)으로 그리고 구체적인 실행은 상향식(bottom-up)으로 진행하는 것이 바람직하다. 특히 시스템구현에 있어서 제한된 규모와 이해관계자들이 포함된 선도프로그램의 실행을 통해서 성공적으로 추진하고 이를 장기적으로 확산시키는 전략이 필요하다.

본 연구는 문헌을 통한 조사를 통하여 이루어졌으며 범위도 계획을 중심으로 한 정책분야에 국한시켰다. 따라서 실제적인 상황 및 실적 등에 대한 조사 등이 부족하여 이러한 계획들이 구체적으로 어떻게 진행되고 있는지에 대한 조사가 미비하다. 따라서 이를 위한 추가 연구가 필요할 것으로 본다.

참 고 문 헌

[1] ANDS(2008), Australian National Data Service Interim Business Plan, http://ands.org.au/andsinterimbusinessplan-final.pdf

[2] ARL(2009), The Research Library's Role in Digital Repository Services - Final Report of the ARL Digital Repository Issues Task Force,http://www.driver-repository.e/media/cs/repository-services-report.pdf

[3] Bell, G., T. Hey, and A. Szalay(2009), Beyond the Data Deluge, Science, Vol. 323, No. 5919, pp.1297-1298.

[4] Francis, R.(2009), Australian eResearch Infrastructure Building with NCRIS and EIF,www.oaklaw.qut.edu.au/files/Francis.ppt

[5] Gray, J, D. Liu, M. Nieto-Santiseteban, D. Dewitt, and G. Heber(2005), Scientific Data Management in the Coming Decade, MSR-TR-2005-10, ACM SIGMOD Record, V 34.4, pp.35-41.

[6] JISC(2004), The Data Deluge: Preparing for the explosion in data, http://www.jisc.ac.uk/media/documents/publications/datadelugebp.pdf

[7] Li, J, W. Li, B. Yan, H. Zhang(2010), Scientific data access activities and infrastructure in China, http://www.codata.org/10Conf/abstracts-presentations/Sessions%20G/G1/Jianhui%20Li.pdf

[8] NSB(2005), Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century, NSB-05-40, http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf

[9] NSTC-IWGDD(2009), Harnessing the Power of Digital Data for Science and Society,

http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

[10] The ANDS Technical Working Group. (2007), Towards the Australian Data Commons, A proposal for an Australian National Data Service, http://ariel.adgrp.com/~ghb/talks/20090202-Net-at-EDU/Meta-University-Simple/thoughtData/430/TowardstheAustralianDataCommons.pdf.

[11] Treloar A.(2008), Design and Implementation of the Australian National Data Service, http://pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf

[12] UK DCC(2008), A Comparative study of international approaches to enabling the sharing of research data, CSD2646, http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/Data-Sharing-Report.pdf.

박 동 진

1983년 아주대학교 산업공학 (공학사)
1988년 한국외국대학교 경영정보학(경영학석사)
1994년 아주대학교 경영대학 경영정보학전공(경영학박사)
1988~1990년 한국생산성분부 정보화사업부 선임연구원
1995년 3월 남서울대학교 경영학과 조교수
1998년 3월~현재 공주대학교 산업시스템공학과 교수
관심분야: 과학데이터관리 및 시각화, 시맨틱웹, 소셜네트워크서비스, 지식경영, 제조정보시스템(ERP, MES)
E-Mail: mispdj@kongju.ac.kr