

상이한 데이터 구조의 데이터베이스간 통합 운영방안 연구*

- 기초학문자료센터를 중심으로 -

A Study of the Integrated Operation for Databases with Different Data Structures

고 영 만(Young-Man Ko)**

배 경 재(Kyung-Jae Bae)***

목 차

- | | |
|----------------------------|------------------------|
| 1. 서 론 | 3.2 토대기초연구 연구성과물 DB 분석 |
| 1.1 연구의 필요성 및 목적 | 4. 통합 모형 설계 |
| 1.2 연구 방법 및 연구 내용 | 4.1 기초학문자료센터 연계 방안 |
| 2. 이론적 배경 | 4.2 데이터 통합 모형 |
| 2.1 이형 데이터 구조의 DB 통합 운영 방안 | 4.3 시스템 구성 방안 |
| 2.2 이형 메타데이터의 통합 운영 방안 | 4.4 방안별 장단점 분석 |
| 3. 통합 대상 시스템 분석 | 5. 결 론 |
| 3.1 기초학문자료센터 시스템 분석 | |

초 록

본 연구에서는 이형 데이터 구조를 가지는 데이터베이스의 통합 운영 가능성을 검토하고 사례 연구를 통한 실제적인 방안을 설계하기 위해서 한국연구재단 토대기초지원사업의 DB 구축 연구성과물과 한국연구재단의 기초학문자료센터 DB 통합 운영 방안을 제시하였다. 토대기초연구지원사업의 성과확산을 위해서는 기초학문자료센터와의 통합 연계가 필수적으로 필요하며, 그 방안으로서 기존에 구축된 DB구축 과제의 DB는 표준 지침을 활용한 XML 데이터베이스화를 필수적으로 진행하되, 향후 구축될 DB구축 과제의 DB는 기초학문자료센터와 연계 시스템을 구성하거나 독립 시스템을 구성하는 방안을 제안하였다.

ABSTRACT

This study reviewed theories for database integration, which combines heterogeneous data structures, and suggested a practical method to integrate databases of Korean Research Memory(KRM) and Infrastructural Basic Research(IBR) as a case study. In order to broadly distribute the outcomes of IBR, it is essential to be connected to and integrated with the database of KRM. As a solution, it was suggested that the current database of IBR should follow standard guidelines as a XML database, and its future database should be integrated with the database of KRM or be established as a stand-alone system.

키워드: 기초학문자료센터, 토대기초연구, 이형데이터 통합, 데이터 매핑, 크로스워크, 메타데이터 레지스트리
Korean Research Memory, Infrastructural Basic Research, Heterogeneous Data Integration,
Data Mapping, Crosswalk, Metadata Registry

* 본 논문은 한국연구재단의 “토대기초연구지원사업 성과확산 강화방안 연구(정책연구-2010-026-인문사회)”의 일부분을 기반으로 함.

본 논문은 2011년도 6월 한국문헌정보학회 춘계학술대회에서 발표한 내용을 수정·보완한 것임.

** 성균관대학교 문과대학 문헌정보학과 교수(ymko@skku.ac.kr) (제1저자)

*** 동덕여자대학교 사회대학 문헌정보학과 전임강사(kjbae@dongduk.ac.kr) (교신저자)

논문접수일자: 2011년 7월 13일 최초심사일자: 2011년 7월 20일 게재확정일자: 2011년 8월 8일

한국문헌정보학회지, 45(3): 69-85, 2011. [http://dx.doi.org/10.4275/KSLIS.2011.45.3.069]

1. 서론

1.1 연구의 필요성 및 목적

온라인 정보 소스와 정보 사용자의 수가 인터넷과 통신 기술의 확산으로 급속히 증가하고 있으며, 이는 정보시스템간의 통합을 요구한다(권도훈 외 2001). 또한 개방과 공유, 참여를 기본 정신으로 삼는 최근의 웹 서비스는 유용한 데이터를 제공하는 다양한 공공 및 민간기관들로 하여금 자관의 데이터베이스를 공개하도록 유도하고 있으며, 이를 통해 양질의 데이터베이스를 생산하고 데이터베이스 운영 관리를 효율화하기 위한 시도로서 서로 다른 형태의 데이터베이스를 물리적으로 교환하고 통합하려는 노력이 계속되고 있다.

특히 학술분야에서 연구성과물의 공유와 확산을 위해 상이한 데이터 구조를 가지는 데이터베이스 간의 상호운용성 확보에 대한 사회적 요구가 높아지고 있고, 이러한 지적산출물의 상이한 데이터 구조와 특성을 고려하여 통합적인 데이터베이스를 생성하여 운영하려는 시도가 여러 기관들을 통해 이루어지고 있다. 이에 따라 다양한 상황에서 이형의 데이터베이스를 통합하기 위한 기술적 측면의 가능성과 실제적인 방안을 연구하는 것이 중요하게 부각되고 있는 실정이다.

본 연구에서는 특정 사례 연구를 통해 이형 데이터 구조를 가지는 데이터베이스의 통합 운영 가능성을 검토하고자 한다. 사례 연구로서는 한국연구재단의 토대기초지원사업의 DB 구축 연구성과물 데이터베이스와 한국연구재단의 기초학문자료센터 데이터베이스의 통합 사

례를 대상으로 한다. 한국연구재단의 토대기초연구지원사업은 인문사회분야의 연구 기초가 되는 핵심 기반 콘텐츠를 마련하기 위한 사업으로서, 인문사회분야의 기초가 되는 연구를 지원함으로써 창조적인 지식생산기반을 구축하기 위하여 동서양 고전 주해·해설, 학술전문·인명사전 편찬, 사회 기초자료 조사 및 통계, 당대사 자료 수집 정리, 역사·민속 지리지, 역사·문화지도 편찬, 희귀 언어 연구 및 기타 토대연구를 지원하고 있다. 특히 인문사회분야 연구의 토대 자료를 정리, 가공함으로써 후속 연구를 위한 기초 자료로 활용하고, 이러한 연구 성과의 확산을 통해 창의적 우수성과물 발굴 강화 및 기초연구의 저변확대를 목적으로 하는 사업이므로 연구성과 확산을 위한 체계 마련이 중요하다.

이와 같은 토대기초연구지원사업의 성과확산을 강화하고, 토대기초연구지원사업 데이터베이스가 각 관련 분야 연구의 기초자료로 활용될 수 있는 실제적인 방법으로서 기초학문자료센터(Korean Research Memory) 데이터베이스와 연계방안을 모색하는 것이 본 연구의 실제적인 연구내용이라고 할 수 있다. 따라서 본 연구는 상이한 데이터 구조를 가지는 데이터베이스의 상호운용성을 확보하는 방안을 특정 사례 연구를 통해 고찰해보고, 가능한 통합 방안을 설계하는 것을 목적으로 한다.

1.2 연구 방법 및 연구 내용

본 연구는 이론적 검토, 시스템 분석, 시스템 설계의 세 가지 방법과 절차를 거쳐 수행되었다.

(1) 이론적 배경 고찰

데이터베이스의 통합은 상이한 데이터구조를 가진 데이터베이스의 통합방안과 상이한 메타데이터의 상호운용성 확보방안의 두 가지 관점에서 고려해 볼 수 있다. 본 연구에서는 각 방안에서 실제로 적용 가능한 방안과 사례를 이론적으로 고찰하였다.

(2) 통합 대상 시스템 분석

본 연구의 사례연구를 수행하기 위해 대상 시스템을 분석하는 절차이다. 시스템을 분석하기 위해 아래 4가지 요소의 분석과정을 수행하였다.

- ① 한국연구재단 토대기초지원사업의 연구성과물 유형 분석
- ② 토대기초지원사업 DB 구축과제의 연구성과로 구축된 DB의 데이터 구조 분석
- ③ 한국연구재단 기초학문자료센터의 데이터 구조 분석
- ④ 한국연구재단 기초학문자료센터의 메타데이터 유형 분석

(3) 통합 모형 설계

시스템 분석 결과를 토대로 아래와 같은 통합 모형을 설계하였다.

- ① 기초학문자료센터와 토대기초지원사업 DB 구축과제의 연구성과물 DB의 통합운영 모형 설계
- ② 기초학문자료센터와 토대기초지원사업 연구성과물의 메타데이터 표준 모형 설계
- ③ 기초학문자료센터 오픈아카이브(Open-Archive) 활용 모형 설계

2. 이론적 배경

2.1 이형 데이터 구조의 DB 통합 운영 방안

이형 데이터 구조의 통합은 전체적인 스키마의 통합과 서로 다른 뷰(View)를 통한 통합으로 구분할 수 있다. 그 중 전체적인 스키마의 통합은 강결합(Tightly-Coupled)과 약결합(Loosely-Coupled) 방식으로 구분할 수 있는데 강결합(Tightly-Coupled)은 통합 데이터 모델을 중심으로 스키마를 설계하여 모든 메타데이터 요소의 링크를 제공하는 방식이며, 약결합(Loosely-Coupled)은 단일 스키마를 제공하기보다 여러 응용방식에 따라 필요한 요구사항에 대해 구조적 통합을 제공하는 개념적 방식이다.

Rakesh Agrawal(1996)은 강결합(Tightly-Coupled) 방식을 통하여 관계형 데이터베이스 시스템에서의 데이터 마이닝 어플리케이션을 개발하였으며, Mauro Sousa(1998)는 병렬 데이터베이스 서버 환경에서의 강결합(Tightly-Coupled) 시스템을 구현하였다. 두 연구 모두 동일한 데이터베이스 환경에서 데이터마이닝 어플리케이션을 강결합(Tightly-Coupled) 방식을 활용하여 이형 데이터 구조의 복잡도를 단순화시키고 텍스트 데이터 마이닝의 성능을 향상시켰다.

W. Bruce Croft(1992)는 인쿼리(Inquiry) 시스템을 통해 객체지향 데이터베이스 시스템에서 텍스트 검색을 위한 구현방식으로 약결합(Loosely-Coupled) 방식을 사용하였는데, IR(Information Retrieval)을 위한 기본 방식으로 공통의 요소에 속하는 정보의 단위(객체

의 단위)를 분리하여 이들 간의 공통성을 학습시키고 이외의 정보 단위들을 대상으로 추출하는 방식을 적용한 바 있다(〈그림 1〉 참조).

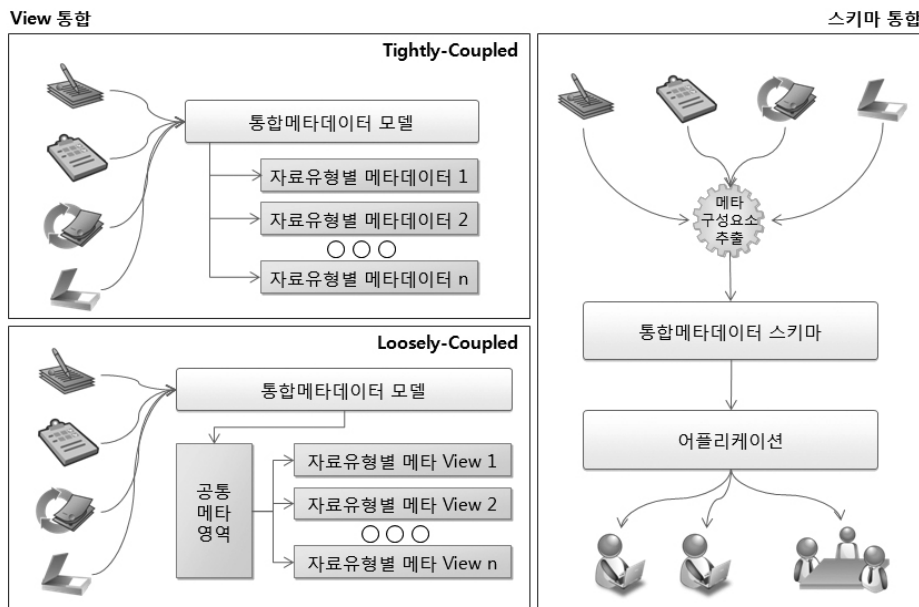
2.2 이형 메타데이터의 통합 운영 방안

연구성과물의 상호운영성을 확보하기 위해서는 연구성과물 유형의 다양성을 인정하고 메타데이터 레벨에서의 상호운용성을 제고하면서 연구성과정보의 통합을 고려해야 한다(박동진 외 2006). 메타데이터의 상호운영을 확보하기 위한 대표적인 방법으로는 크게 메타데이터 크로스워크를 만들어 사용하는 방법, 자원의 속성을 감안하여 다양한 메타데이터 형식과 기술구조를 인정하고 상호 매핑을 통해 해결하는 방법, 그리고 메타데이터 레지스트리(Metadata Registry: MDR)에 의한 해결 방법의 세 가지를

들 수 있다(고영만, 서태설 2007).

2.2.1 메타데이터 매핑

시스템의 기능 수정을 최소화하면서 안정적으로 기록물을 입수할 수 있도록 하기 위해 고안된 기능이 '메타데이터 매핑 도구(Metadata Mapping Organizer)'이다(박석훈 2008). 메타데이터 매핑은 정보자원의 특성을 감안하여 다양한 메타데이터 형식과 기술구조를 인정하고 이들 간의 상호 운용을 위한 방법으로, 상호 참조 테이블 방식과 범용 메타데이터 통합방식이 존재한다. 이형 메타모델간의 요소(Element) 레벨에서 동질성을 지니는 요소를 연결하여 이들 간의 매핑을 진행하는 방법으로 메타데이터 매핑의 목적에 따라 동질성 참조 요소가 변경 될 수 있으며 설계자의 주관적인 의지가 반영될 수 있는 단점이 있다. 사례



〈그림 1〉 이형 데이터 구조의 통합운영방안

로써 대통령기록영구관리시스템에 도입되어 있는 “메타데이터 매핑 도구(Metadata Mapping Organizer)”는 대통령비서실기록관리 시스템, 온나라시스템, 자료관시스템, 전자문서시스템, 행정정보시스템, e지원시스템등 다양한 시스템에 산재되어 있는 다양한 포맷 유형의 전자기록물을 안정적으로 이관 받고 관리 할 수 있는 기능을 구현하였다. 위 시스템에서 메타데이터의 매핑은 인수된 기록물에 대해서 먼저 파일과 메타데이터 구조를 분석하고 분석된 결과를 바탕으로 대통령기록물영구관리시스템의 ‘필수 메타 데이터’ 해당 항목과 연결매핑을 하는 ‘매핑규칙’ 정의를 진행한 후 정의된 매핑규칙에 따라 자동으로 분류를 진행한다.

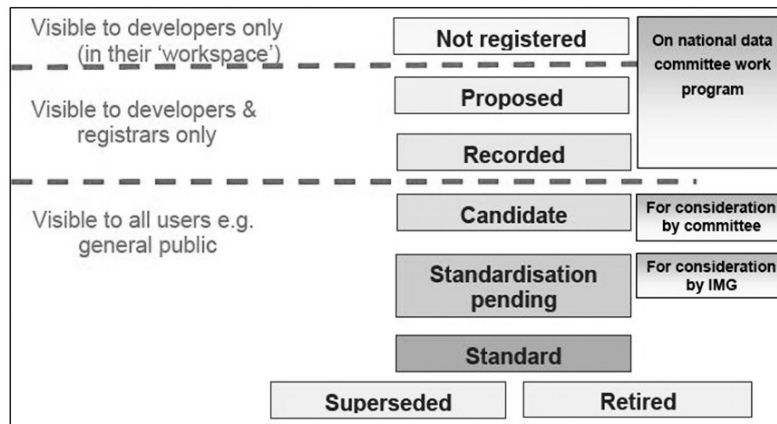
2.2.2 크로스워크

크로스워크는 하나의 메타데이터 스킴을 다른 메타데이터 스킴으로 요소, 의미, 구문을 매핑하는 방법이다(NISO 2004). 레코드의 상호교환이나 종합 목록에 대한 제공 및 메타데이터 수집(Harvesting)을 위해 어느 한 메타데이터 스킴으로부터 다른 스킴으로 레코드를 물리적으로 변환하기 위한 기본 규격으로 사용하며, 검색 엔진에 의해 서로 다른 데이터베이스에서 동일하거나 유사한 내용을 가진 필드들에 질의하기 위해 사용된다. 특징으로는 익숙하지 않은 스킴을 접할 때 더 친숙한 스킴 사이의 크로스워크를 대신 이용하는 경향이 있기 때문에 메타데이터 요소들의 의미와 용도 추론에 한계가 존재하며, 상대적으로 전문가의 지적인 능력이 필요하고, 잦은 수정에 대한 유지관리 부담이 발생하며, 메타데이터 스킴 사이의 정확한 1:1 대응이 이루어지기 어려운 측면이 있다(Chan

and Zeng 2006; Zeng and Chan 2006). 이러한 부분을 기계적으로 처리하기 위한 노력으로 OCLC는 메타데이터간 크로스워크를 실현한 API인 크로스워크 웹서비스 API(Crosswalk Web Service API)를 공개했다. 이 API는 메타데이터를 표준(MARC, Dublin Core, MODS 등)과 포맷(XML, RDF, ISO 2709 등), 그리고 문자셋(MARC8, UTF-8, Windows 1251 등)으로 규정하여 정의하고 있으며 해당 API에 파라미터로써 특정 값(문자열)을 넘기면 정의된 규칙에 따라 매핑된 결과 값(문자열)을 받을 수 있다. 정의된 규칙은 WSDL(Web Services Description Language)로 OCLC 크로스워크 웹서비스 데모 웹사이트에서 확인할 수 있다.

2.2.3 메타데이터 레지스트리

메타데이터 레지스트리는 다양한 종류의 메타데이터 표준을 메타데이터 요소의 설명을 통하여 메타데이터의 공유를 가능하게 한다(이태웅 외 2002). 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지관리하며 메타데이터 요소간의 호환성을 유지하기 위한 표준 규격으로 메타데이터 요소의 명세와 표준화를 위해 필수적인 데이터 요청과 등록을 위한 모델을 제공한다. ISO/IEC 11179는 ISO/IEC JTC1 SC32의 WG2에서 개발한 메타데이터 레지스트리에 관한 국제표준이다. 이 표준에서는 메타데이터 레지스트리의 구축에 관한 구체적인 시스템 스펙을 제공하고 있지는 않지만 개념 및 논리 데이터 모델을 제시하고 있고 이에 따라 총 6가지의 세부규격을 제공한다. MDR은 상호운용성을 보장하고 이를 관리하기 위하여 의미, 구문,



〈그림 2〉 METeOR의 메타데이터의 등록단계

표현을 표준화하고 메타데이터의 명세와 의미 공유를 위한 다양한 정의를 제공한다. 또한 이러한 정의를 기반으로 시스템 설계 시 표준화된 모델링 및 스키마 설계에 대한 가이드를 제공하고 있다.

대표적인 사례로써 미국 국립암연구소(NCI: National Cancer Institute)의 암 표준데이터 저장소 시스템인 caDSR(Cancer Data Standards Repository)은 NCI의 하위 기관인 caCORE(Cancer Common Ontologic Representation Environment)의 인프라로 구축된 시스템이다. 또다른 사례로는 AIHW(Australian Institute of Health and Welfare)에서 구축한 METeOR(Metadata Online Registry) 시스템으로서 건강(Health), 커뮤니티 서비스(Community services), 주거 지원(Housing assistance) 세 분야에 대한 국가 메타데이터 표준을 제공하고 있다(〈그림 2〉 참조).

3. 통합 대상 시스템 분석

3.1 기초학문자료센터 시스템 분석

3.1.1 시스템 개요

한국연구재단의 기초학문자료센터(Korean Research Memory: KRM)는 재단의 학술연구조성사업에 의해 지원된 인문·사회과학·예술분야의 연구성과물을 데이터베이스화하여 대학원생 이상의 전문 연구자를 대상으로 서비스하는 데이터베이스 시스템이다.

학술연구조성사업에 의해 지원된 연구과제의 중간산출물 및 연구과정에서 수집했거나 생성한 원자료 등을 체계적으로 관리하고 공유하는 시스템의 필요성이 점차 대두되고, 국가적 차원의 연구산출물 및 관련 기초자료의 디지털 아카이빙 시스템 구축을 위한 표준화 지침이 필요하게 되면서 기초학문자료센터 구축의 필요성이 제기되었다. 또한 기초학문분야 전문 연구자에게 학술적 가치가 높은 일차자료(원자료, 사실정보 등)를 제공하여 후속연구의 활성화

화를 유도한다는 측면에서 기초학문자료센터의 유용성이 인정받고 있다. 기초학문자료센터 구축 대상 자료로서는 원자료, 연구결과물, 중간 산출물이 해당된다(임건면 외 2007).

3.1.2 시스템 구성현황

기초학문자료센터 시스템은 크게 연구성과물 DB 구축 시스템, 오픈아카이브 시스템, 용어 데이터레지스트리의 3개의 하부 시스템으로 구성된다.

(1) 연구성과물 DB 구축 시스템

재단의 연구과제는 하나의 과제로부터 여러 유형의 연구성과물이 생성되는 경우가 빈번하고, 산출되는 파일 종류와 개수가 다양하다. 이에 기초학문자료센터 시스템에서는 다양한 유형과 매체를 가지는 연구성과물의 복잡한 구조를 제대로 기술하고자 계층적 네트워크 구조를 제시하는 FRBR(Functional Requirements for Bibliographic Records) 개념모형을 도입하였다. 따라서 기초학문자료센터 데이터 구조는 논리적 수준의 계층구조 목록을 위하여 데이터베이스 구축대상인 연구성과물 정보를 FRBR 모형의 제1집단에서 제시하는 저작, 표현체, 실현체, 개별자료의 4가지 개체로 정의하고 있다.

- 저작: 재단 또는 외부에서 지원한 연구과제
- 표현체: 연구과제를 수행하며 생산한 연구 성과물
- 실현체: 연구성과물이 디지털매체로 변환·저장된 파일
- 개별자료: 연구성과물의 실물자료에 대한 개별 객체

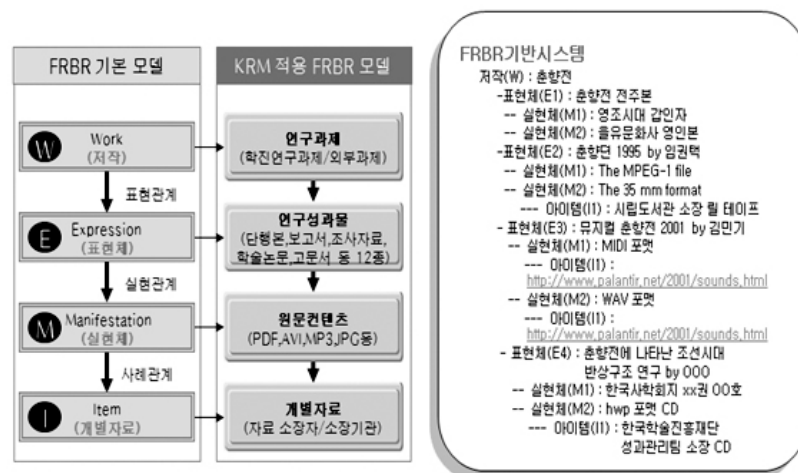
연결 구조를 지닌 목록을 위하여 연구과제를 최상위 계층인 '저작'으로 두고, '저작'인 연구과제의 고유한 식별번호를 연결장치로 사용하여 연구성과물의 상·하위 계층과 동일 계층 간에 연결 구조를 지니고 있다. 연계 가능한 모든 정보와의 연결 기능을 위하여 제1집단의 행위자에 속하는 제2집단의 적용으로 '통합연구인력정보 데이터베이스'와 '기관정보 데이터베이스'를 연동하였고, 제1집단과 주제관계에 있는 제3집단의 적용으로 '연구분야분류표'와 '데이터 레지스트리 시스템'을 연동하고 있다(<그림 3> 참조).

(2) 오픈 아카이브 시스템

기초학문자료시스템은 국내외 연구자들의 연구성과를 공유하고 상호 평가하는 커뮤니티 기반의 오픈아카이브 서비스를 제공하고 있다. 이 시스템은 주제 분야별 연구자 그룹 형성을 지원하는 커뮤니티 시스템(학술 커뮤니티)을 포함하고 있으며, 연구자 자신의 저작물을 직접 제출할 수 있는 셀프아카이빙(Self-Archiving) 기능과 RSS 및 OAI-PMH DP(Open Archives Initiative Protocol for Metadata Harvesting - Data Providers)를 통한 외부 학술연구정보 수집 및 제공 기능을 제공하고 있다.

(3) 용어데이터 레지스트리

한국연구재단의 용어구축사업에 의해 구축된 용어를 검색할 수 있는 시스템으로서, 재단 과제계획서 색인어, 기초학문자료센터 연구성과물 DB 구축 시 입력된 비통제 주제어, 재단 분야분류표 용어 등을 기본으로 하여 구축되었다. 2009년 3월 현재, 구축기에 등록된 전체 용



〈그림 3〉 FRBR 개념모형을 적용한 기초학문자료센터 DB 개념도

어 수는 74,573 용어이고, 용어간 관계 설정 수는 168,700 여건이다.

3.1.3 데이터 현황

한국연구재단은 매년 대학 부설 연구소 및 연구기관을 대상으로 DB구축 사업단을 선정하여 연구성과물 DB를 구축하고 있으며, 2010년 5월 현재까지 기초학문자료센터에 구축된 연구성과물DB는 총 343,087건으로, 연구성과물DB 유형별 구축 현황은 〈표 1〉과 같다.

3.2 토대기초연구 연구성과물 DB 분석

3.2.1 연구성과물 DB 유형 분석

2005년부터 2010년까지 신청 과제 총 919건 (선정과제 171건)에 대해 인문학분야 중장기 토대연구 지원방안 연구(조광 2009)에서 제시한 연구성과물 유형 분류 기준을 적용하여 아래와 같이 총 5가지 연구 유형으로 분류될 수 있다.

- 사전학(백과, 전문, 목록, 색인 등)
- 번역학(국학, 동양, 서양 등)
- 자료학(정본화, 해제, 현지조사, 字學, 구술사, 전자문화지도 등)
- 총서학(대관, 학술사, 연구사, 레퍼런스 북 등)
- 기타(여성학, 생태인문학 등)

〈표 1〉 기초학문자료센터 연구성과물 DB 유형별 구축 현황

(2010년 5월 기준, 단위: 건)

단행본	보고서	조사자료	학술논문	고문서	고도서	이미지	동영상	녹음자료	웹사이트	기사자료	날장자료
2,525	5,427	2,630	14,984	12,999	3,322	56,888	5,410	213,312	1,214	3,726	20,650

〈표 2〉 토대기초연구지원사업 연도별 DB 구축과제 현황

(2011년 2월 기준, 단위: 건)

구 분	2005년	2006년	2007년	2008년	합계
결과보고 완료과제	51	17	27	9	104
DB 구축과제 (구축율)	20 (39.2%)	9 (52.9%)	17 (63.0%)	3 (33.3%)	49 (47.1%)

* 2009년, 2010년 선정과제의 경우, 결과보고 완료 현황(총 1건)이 미비하여 해당 분석에서 제외함.

3.2.2 DB구축 과제 현황

2005년~2008년(선정연도 기준) 토대기초연구지원사업 연도별 DB 구축과제 현황은 〈표 2〉와 같다. 2005년~2008년에 선정된 연구과제 중 결과보고 완료과제에 대해 DB 구축 현황을 조사한 결과, 총 104건의 완료과제 중 49건(47.1%)이 DB를 구축한 과제로 나타났으며, 선정연도별 DB 구축율은 2005년 39.2%, 2006년 52.9%, 2007년 63.0%, 2008년 33.3%로 파악되었다. 거의 절반에 가까운 비율의 토대기초연구사업 연구 과제가 DB 구축형 과제이므로 해당 사업의 사회적 활용성을 향상시키기 위해서 기초학문자료센터의 연구성과물 DB 구축 시스템과의 연계 통합 방안을 설계할 필요가 있다.

4. 통합 모형 설계

기초학문자료센터 데이터베이스와 토대기초연구지원사업 연구성과물 데이터베이스의 통합운영 모형을 설계하기 위해서는 향후의 각 시스템 연계 통합 방안과 기존 데이터의 통합 방안 두 가지 측면이 모두 고려되어야 한다. 향후의 각 시스템 연계 통합방안은 기초학문자료센터 시스템을 활용하여 연계하는 방안과 함께 각각

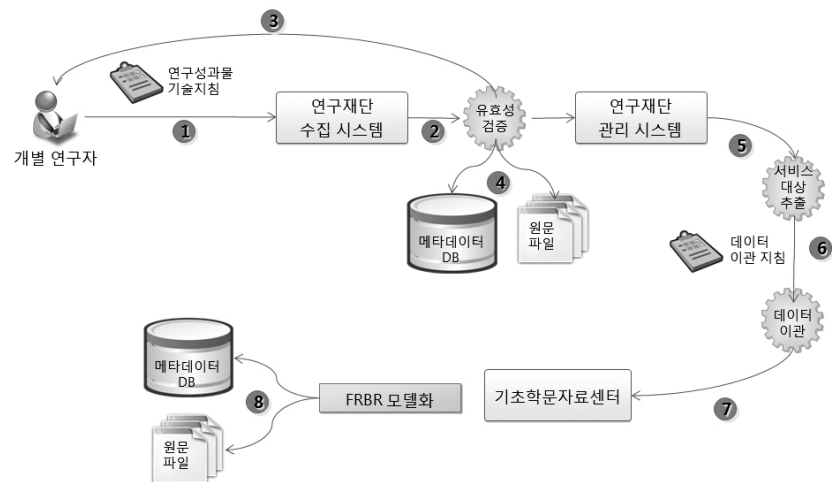
독립 시스템을 유지하여 활용하는 방안 두 가지가 설계되었다.

4.1 기초학문자료센터 연계 방안

4.1.1 기초학문자료시스템 활용방안

토대기초연구지원사업의 연구성과물을 지속적으로 유지·관리하며 안정적인 서비스를 제공하기 위해 기초학문자료센터를 통해 수집된 데이터를 서비스하는 모델이 필요하게 되었고 이에 대한 방안으로 강결합(Tightly-Coupled) 방식의 표준 XML 문서에 기반한 통합모형을 설계하였다. 〈그림 4〉는 구체적인 연계 프로세스를 정의한 것이다.

- ① 개별연구자 혹은 사업단에서 연구성과물 기술 지침을 토대로 연구재단 수집 시스템을 통하여 연구성과물을 저장함
- ② 수집된 데이터는 유효성 검증 과정을 진행함
- ③ 유효성이 입증되지 않은 성과물은 원 저작자에게 보완을 요청함
- ④ 검증단계를 통해 한국연구재단의 메타데이터 DB와 원문파일 저장소에 저장이 되고 이후 재단 관리시스템에 등록이 처리됨



〈그림 4〉 연구성과물의 기초학문자료센터 연계 프로세스

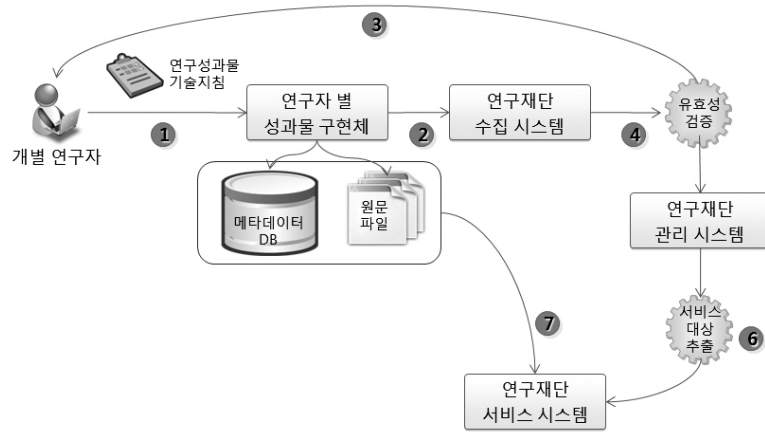
- ⑤ 이후 기초학문자료센터를 통해 서비스 될 연구 성과물을 추출함
- ⑥ 추출된 데이터는 데이터 이관 지침을 통해 기초학문자료센터로 이관 대기상태로 변환
- ⑦ 기초학문자료센터에서 이관 대기상태의 데이터를 이관 받아 내부 시스템에 적재함
- ⑧ 서비스를 위해 FRBR 데이터 유형으로 모델링을 진행함

4.1.2 독립시스템 운영방안

독립시스템 운영방안은 OAI-PMH에 기반한 셀프아카이빙(Self-Archiving) 시스템으로 설계되었다. 이 시스템은 기초학문자료센터의 연구성과물 DB 구축 시스템을 활용하지 않고 토대기초연구지원사업의 연구성과물 작성 주체가 제공하고자 하는 데이터에 적합한 시스템을 구성하여 연구자의 연구 목적에 맞는 연구성과물 서비스가 가능한 모델이다. 이 방안은 연구 기간 동안 혹은 연구 완료 이후에도 관리의 주

체가 해당 연구자가 되므로 지속적인 데이터의 업데이트를 제공할 수 있는 장점이 있는 반면 개별 연구자가 성과물의 활용 및 관리를 지속해야 한다는 단점이 존재한다. 〈그림 5〉는 구체적인 연계 프로세스를 정의한 것이다.

- ① 개별연구자 혹은 사업단에서 연구성과물 기술 지침을 토대로 연구자별 성과물 구현체 생성
- ② 연구재단 수집 시스템을 통해 등록
- ③ 수집된 데이터는 유효성 검증 과정을 진행함
- ④ 유효성이 입증되지 않은 성과물은 원 저작자에게 보완을 요청함
- ⑤ 연구재단 관리시스템을 통해 데이터 등록 진행
- ⑥ 외부에 서비스 될 연구 성과물을 추출함
- ⑦ 연구재단 서비스 시스템에서 연구자별 성과물 구현체에 대한 서비스 제공



〈그림 5〉 연구성과물의 독립 운영을 위한 프로세스

4.2 데이터 통합 모형

4.2.1 기존 데이터 수집 방안

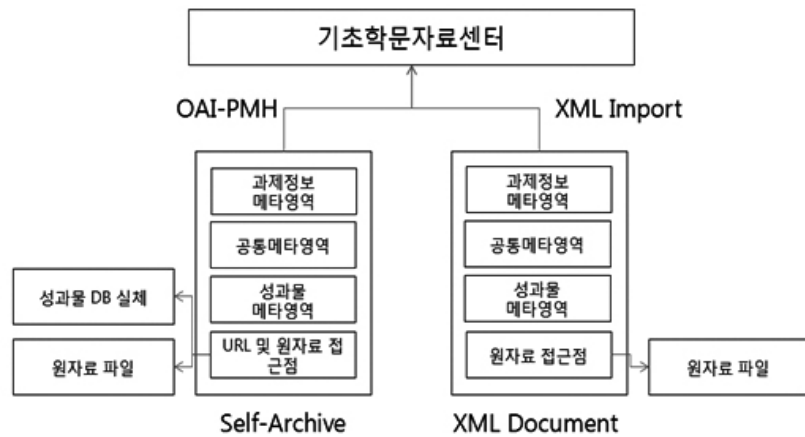
기존에 구축된 토대기초연구사업의 연구성과물 데이터를 기초학문자료시스템에 적합한 형태로 수집하여 통합하는 방안은 XML 데이터베이스화와 셀프 아카이빙(Self-Archiving)의 두 가지 방식으로 고려되었다.

연구 성과물의 XML 데이터베이스화는 연구성과물(연차보고서, 결과보고서, 최종결과물)과 함께 구축된 DB 성과물을 DB구축 과제를 위한 표준 XML 스키마를 적용토록 하여 XML 문서와 원 자료를 함께 제출하는 방안으로서, 한국연구재단에서 제공하는 표준 지침을 활용하여 XML 문서를 구축토록 한다. 구축된 연구성과물 DB는 성과물 등록시스템에서 제공하는 반입도구를 통하여 성과물 제출 시스템에 등록되는 과정이 진행된다.

셀프 아카이빙(Self-Archiving)은 디지털 자료의 메타데이터와 원본데이터를 웹 기반에서 접근이 가능하도록 저장하는 것을 지칭하며 자

료 작성자 스스로 데이터 구축을 진행하는 연구성과물 제출 유형이다. 자체적으로 구축된 DB를 서비스하거나 별도의 DB 구축 시스템을 활용하는 과제를 위해 셀프 아카이빙(Self-Archiving) 시스템을 제공하여 연구과제에 적합한 DB 구축을 지원하는 방안이라고 할 수 있다. 이 방안은 별도의 DB 제출과정을 거치지 않고 연구성과물 등록 시 해당 시스템에 접근가능 한 OAI-PMH 서비스 URL을 등록하는 것으로 DB 제출이 완료되는 형태이다.

셀프 아카이빙(Self-Archiving)의 경우, 해당 데이터에 접근하기 위한 URL과 원자료에 접근하기 위한 접근점(URI)을 명시해야 한다. 이와 같이 XML 데이터베이스화의 경우에도 함께 제출된 원자료 파일에 접근할 수 있는 접근경로가 제공되어야 한다. 셀프 아카이빙(Self-Archiving)과 XML 문서 모두 최종적으로 기초학문자료센터에 연계되는 형태는 XML 구조이다(〈그림 6〉 참조).



〈그림 6〉 DB 구축 연구성과물의 기초학문자료센터 연계 방안

4.2.2 메타 영역별 구조

본 연구에서는 각기 다른 데이터 유형을 과제관리를 위한 공통요소와 토대분야 연구가 지니고 있는 공통요소들을 추출하여 이들을 각각 과제정보 메타영역, 공통 메타영역, 성과물 메타영역으로 분리하고 과제정보 메타영역과 공통 메타영역에 해당하는 정보들은 통합검색 시스템에 적용하여 과제 관리를 위한 어플리케이션에 활용되도록 설계하였다. 이는 이형 데이터 구조의 통합 운영을 위해 약결합(Loosely-Coupled) 방식을 적용한 것이다. 또한 고급 연구자 및 관련분야 연구자들을 위해 상세 검색을 제공하는 연구과제 검색시스템을 통해서 연구성과물의 상세영역인 성과물 메타영역의 데이터를 활용하도록 설계하였다(〈그림 7〉 참조).

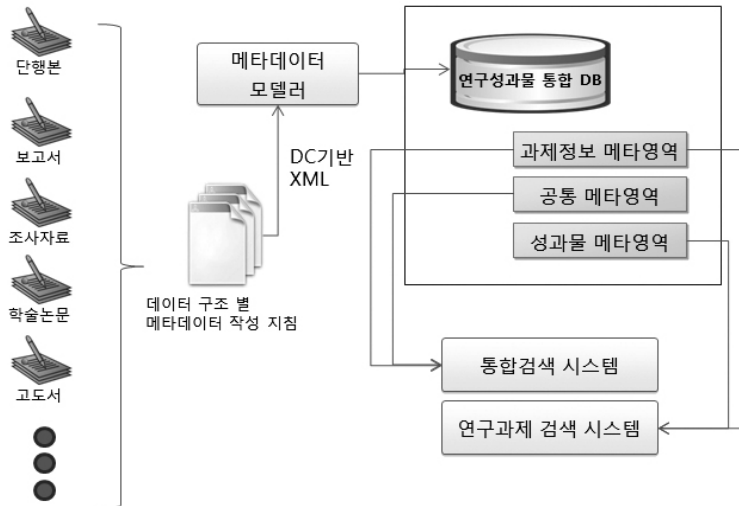
과제정보 메타영역은 해당 연구과제의 일반적인 내용을 기술하는 부분으로 연구과제번호, 연구과제명, 연차표기 등이 포함된다. 공통메타영역은 일반적인 메타데이터가 지니는 속성을 기술하는 것으로 자료명, 저자명, 초록, 목차 등이 포함된다. 성과물 메타영역은 해당 데이터의

특성이 기술된 부분으로 단행본의 경우 발행일, ISBN 등이 포함되며 이미지의 경우 촬영정보, 촬영장소, 피사체 배경정보 등이 해당된다.

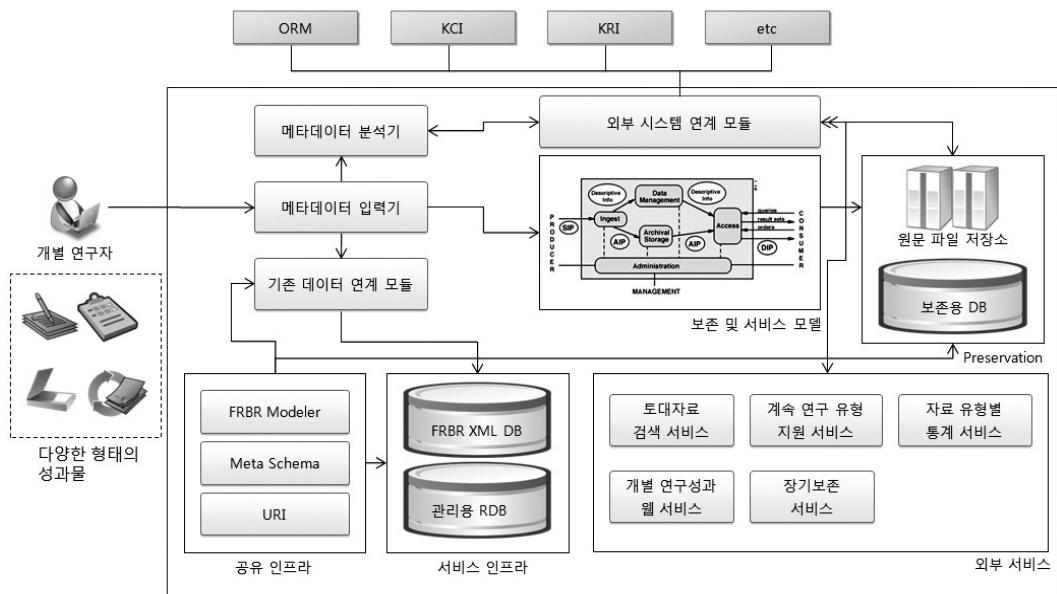
4.3 시스템 구성 방안

위에서 언급한 2가지 방식으로 기초학문자료센터로 입수된 토대기초연구 연구성과물 데이터는 〈그림 8〉과 같은 통합된 시스템 환경에 따라 각각의 단계를 거쳐 서비스되어지며 각 단계는 다음과 같다.

- 메타데이터 입력기: 연구성과물의 데이터 정보를 입력하기 위해 각 성과물의 유형에 적합한 메타데이터 입력기를 통해 초기 데이터를 생성함
- 메타데이터 분석기: 입력된 메타데이터의 속성을 분석하여 타 시스템과 연계할 항목을 도출하고 데이터 이관을 위한 메타데이터 패키지를 생성함(분석의 항목으로는 외부 공개여부나 저작권사항, 중복체크 등이 포함됨)



〈그림 7〉 약결합(Loosely-Coupled) 방식의 적용 모형



〈그림 8〉 통합시스템 모형도

- 기존 데이터 연계 모듈: 기초학문자료센터에서 서비스 되고 있는 기존 서비스 모델에 연계하기 위해 데이터를 분해하는

과정으로 입력된 메타요소들이 서비스 인프라에 적합하도록 정제하는 과정임

- 공유인프라: 등록된 연구성과물 데이터를

한국연구재단내의 타 서비스나 글로벌 유통을 위해 해외 데이터센터등과 연결하기 위해 필요한 기반 시스템

- 외부서비스 연계 모듈: 한국연구재단내에서 서비스 되는 각 시스템들에 토대기초 연구지원사업의 성과물들을 연계하기 위한 내부 시스템을 구축할 필요가 있음
- 보존 및 서비스 모델: 현재 OA 서비스로 제공되고 있는 시스템을 확장하여 제출된 연구성과물의 원자료를 아카이빙 하는 부분으로 장기보존과 데이터 유통을 위한 기본 데이터 모델
- Preservation: 보존 및 서비스 모델로부터 도출되는 연구성과물 원자료에 대하여 장기보존이 가능한 파일형태로 변환하고 이에 적합한 메타데이터를 RDBMS에 저장하는 부분

4.4 방안별 장단점 분석

기초학문자료센터와 토대기초연구 DB를 연계하는 방안은 연구성과물의 장기보존 서비스를 제공하여 일회성 제출로 그치지 않고 지속적인 후속연구 지원에 활용될 수 있다는 장점이 있다. 이런 점은 독립시스템이 지니고 있는 서비스 지속성의 위험요소를 제거할 수 있으며, 기초학문자료센터의 통합 정보서비스에 활용될 수 있어서 기존 연구성과물과의 융합 서비스가 가능할 뿐만 아니라 주제별 검색, 의미기반 검색, 특정분야 연구유형 분석, 최근 연구동향 파악 등 다양한 형태의 연구지원 서비스가 가능하다. 특히 독립 시스템 운영에 발생하는 시스템 구축비용을 절감하여 예산을 효율화할

수 있으며, 지속적인 운영·유지 비용을 최소화할 수 있다. 또한 이 방안은 기초학문자료센터를 통한 연구 성과물 데이터의 체계적 관리를 가능하게 해서 각종 통계 및 서비스 활용 유형 분석 등을 자유롭게 활용할 수 있다. 그러나 이런 장점에 반해 기초학문자료센터가 이런 연계 시스템을 구축하기 위해서는 추가적인 시스템 개발이 필요하며, 각 연구자들의 시스템 활용 교육이 뒷받침 되어야 하고 한편으로는 개별 연구성과물의 특성을 살리지 못하는 획일적인 서비스 구성이 우려될 수 있다.

토대기초연구 독립시스템 운영 방안의 경우에는 제공하고자 하는 데이터에 적합한 고유 시스템을 구성하여 연구자의 연구 목적에 맞는 연구성과물 서비스가 가능하다. 연구 기간 동안 혹은 연구 완료 후에도 관리의 주체가 연구자가 되므로 지속적인 데이터의 업데이트를 제공할 수 있다는 점과 연구 데이터 별 특화된 서비스로 발전시킬 수 있다는 부분이 장점으로 도출될 수 있을 것이다. 그러나 데이터 서비스를 위해 준비되는 리소스가 과다할 수 있다는 점과 별도의 독립 시스템을 개발하기 위해 시스템 개발이 불가피하며 이 시스템 개발을 위해 연구비의 상당한 부분이 소비될 수 있다는 점, 또한 지속적인 유지보수 혹은 업데이트를 위해 연구자가 부담하여야 하는 부하가 예상된다. 따라서 연구기간 동안만 운영되고 이후 지속적으로 유지되지 못할 가능성 또한 높기 때문에 양질의 연구성과물이 후속연구에 활용되지 못할 수 있다는 단점도 동시에 나타난다.

위의 각 방안의 장단점을 종합해볼 때, 토대기초자료 연구성과물을 기초학문자료센터 시스템과 통합 연계하는 방안이 기본적으로 타당

한 것으로 사료된다. 통합 연계 방안은 토대기초자료 연구성과물과 기초학문자료센터의 연구성과물 유형 분석을 통해 일반화될 수 있는 데이터 유형 몇 가지를 정의하고, 해당 데이터 유형 입력 방식에 대한 별도의 지침을 정의하여 연계토록 하는 방식이 적절할 것이다. 한편 연구성과물 고유 성격상 일반화되기 힘든 특별한 형태의 DB 구축 연구과제는 독립서비스 방식을 취하는 것도 가능할 것으로 보인다. 그러나 독립서비스 방식을 취할 경우, 해당 방식의 단점을 보완하기 위해 웹 기반으로 구축된 데이터의 접근 가능성, 기초학문자료센터와의 연계를 위한 OAI-PMH 방식의 데이터 서비스 제공, 최소 일정기간 이상의 서비스 유지 관리 등을 전제조건으로 하여 규약화할 필요가 있다.

5. 결 론

이형 데이터 구조를 가지는 데이터베이스의 통합 운영 가능성을 검토하고 사례 연구를 통한 실제적인 방안을 설계하기 위해서 한국연구재단 토대기초지원사업의 DB 구축 연구성과물과 한국연구재단의 기초학문자료센터 DB 통합 운영 방안을 제시하였다.

현재 토대기초연구지원사업의 DB 구축 연구성과물은 다년간 우수한 인력이 구축한 양질의 자료임에도 불구하고 성과확산을 위해 필요한 관리 방안이 마련되어 있지 않고 있어서 기존에 이미 성공적으로 활용되고 있는 연구성과물 시스템과의 통합 운영방안이 필요한 상황이

다. 특히 동일기관에서 구축하는 DB의 경우 구조적으로 동일한 규칙에 따라 자료의 수집과 DB 구축이 이루어져야 지속적인 관리와 부가적인 서비스 창출이 가능하기 때문이다.

연구성과물의 체계적 관리와 확산을 위해서는 구조적으로 동일한 규칙에 따라 DB 구축이 이루어져야 하며, 따라서 현재 기초학문자료센터에서 활용하고 있는 FRBR 데이터 모델에 따라 관리될 수 있는 데이터 형식이 필요하다. 이에 따라 향후 구축되는 토대기초연구지원사업의 DB는 반드시 표준 지침에 따라 구축되어야 할 것이다.

토대기초연구지원사업의 성과확산을 위해서는 기초학문자료센터와의 통합 연계가 필수적으로 필요하며, 본 연구에서는 그 방안으로서 2011년 이전에 구축된 DB구축 과제의 DB는 표준 지침을 활용한 XML 데이터베이스화를 필수적으로 진행하되, 향후 구축될 DB구축 과제의 DB는 기초학문자료센터와 연계 시스템을 구성하거나 독립 시스템을 구성하는 방안을 제안하였다.

기본적으로 토대기초연구지원사업과 기초학문자료센터와의 통합 연계 방안은 토대기초자료 연구성과물과 기초학문자료센터의 연구성과물 유형 분석을 통해 상호 통합 연계하는 한편 연구성과물 유형의 특수성을 감안할 때 일반화되기 힘든 특별한 형태의 DB 구축 연구과제는 개별적인 독립서비스 방식을 취하도록 보완책을 마련하여 다양한 연구성과물의 품질을 보장하는 것도 가능할 것이다.

참 고 문 헌

- [1] 고영만, 서태설, 임태훈. 2007. 의미 호환을 위한 메타데이터 매핑 연구. 『정보관리학회지』, 24(4): 223-238.
- [2] 권도훈, 박성공, 이정욱, 백두권. 2001. 표준 인터페이스를 이용한 데이터베이스 통합. 『한국정보과학회 봄 학술발표논문집』, 28(1B): 145-147.
- [3] 박동진, 이상태, 최기석. 2006. 메타데이터 기반의 연구성과정보 검색시스템의 개념적 설계. 『정보관리연구』, 37(2): 1-20.
- [4] 박석훈. 2008. 대통령기록관리시스템의 구축. 『기록인』, 2008(봄): 56-61.
- [5] 이태웅, 권주흠, 백두권. 2002. 메타데이터 레지스트리 기반의 XML 스키마 생성기법. 『한국정보과학회 2002년도 가을 학술발표논문집』, 29(2(I)): 181-183.
- [6] 임건면, 노승익. 2007. 『기초학문자료센터 운영을 위한 제도연구 -법률적 문제, 특히 저작권을 중심으로』. [서울]: 한국학술진흥재단.
- [7] 조광. 2007. 『연구소 지원을 통한 인문학 인프라 구축과 전임연구인력 활용 방안에 대한 연구』. [서울]: 한국학술진흥재단.
- [8] 기초학문자료센터. [online]. [cited 2010.7.1]. <<http://www.krm.or.kr>>.
- [9] Chan, L. M., & Zeng, M. L. 2006. "Metadata interoperability and standardization-a study of methodology part I: Achieving interoperability at the schema level." *D-Lib Magazine* [online], 12(6). [cited 2011.8.3]. <<http://www.dlib.org/dlib/june06/chan/06chan.html>>.
- [10] Dublin Core Metadata Initiative. [online]. [cited 2010.7.1]. <<http://dublincore.org>>.
- [11] Sousa, Mauro, Mattoso, Marta, & Ebecken, Nelson F. F. 1998. "Data-Mining: A tightly-coupled implementation on a parallel database server." *International Workshop on Database and Expert Systems Applications*, 9: 711-716.
- [12] NISO (National Information Standards Organization). 2004. *Understanding Metadata*. Bethesda, MD: NISO Press. [online]. [cited 2011.8.3]. <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>.
- [13] OCLC Crosswalk Web Service Demo. [online]. [cited 2010.7.1]. <<http://xwalkdemo.oclc.org/webservices/metadata/crosswalk?wsdl>>.
- [14] Agrawal, Rakesh, & Shim, Kyuseok. 1996. "Developing tightly-coupled data mining applications on a relational database system." *International Conference on Knowledge Discovery and Data Mining*, 2: 287-290.
- [15] Croft, W. Bruce, Smith, Lisa Ann, & Turtle, Howard R. 1992. "A loosely-coupled integration

of a text retrieval system and an object-oriented database system.” *In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 223-232.

- [16] Zeng, M. L., & Chan, L. M. 2006. “Metadata interoperability and standardization-a study of methodology part II: Achieving interoperability at the record and repository levels.” *D-Lib Magazine* [online], 12(6). [cited 2011.8.3].
 <<http://www.dlib.org/dlib/june06/zeng/06zeng.html>>.

• 국문 참고자료의 영어 표기

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

- [1] Ko, Young-Man, Seo, Tae-Sul, & Lim, Tae-Hoon. 2007. “A study on metadata mapping for semantic interoperability.” *Journal of the Korean Society for Information Management*, 24(4): 223-238.
- [2] Kwon, Do-Hoon, Park, Sung-Kong, Lee, Jeong-Oog, & Baik, Doo-Kwon. 2001. “An integration of databases using standardized interfaces.” *Proceedings of the Korea Information Science Society Spring Conference*, 28(1B): 145-147.
- [3] Park, Dong-Jin, Lee, Sang-Tae, & Chio, Ki-Suk. 2006. “Conceptual design of metadata based research results information retrieval system.” *Journal of Information Management*, 37(2): 1-20.
- [4] Park, Suk-Hoon. 2008. “A study on the presidential archives management system.” *Kirokin*, 2008(Spring): 56-61.
- [5] Lee, Tae-Woong, Kweon, Ju-Hum, & Baik, Doo-Kwon. 2002. “XML schema generating technique based on metadata registry.” *Proceedings of the Korea Information Science Society Fall Conference*, 29(2(I)): 181-183.
- [6] Lim, Kun-Myun & Noh, Seung-Ik. 2007. *System Research for Administering Korea Research Memory Center-Legal Issue(Intellectual Property)*. [Seoul]: National Research Foundation of Korea.
- [7] Cho, Kwang. *Studies on the Measures to Support Mid-and Long-Term Fundamental Research in Humanities Field*. Seoul: National Research Foundation of Korea.
- [8] Korean Research Memory. [online]. [cited 2010.7.1]. <<http://www.krm.or.kr>>.

