

# Metadata Registry를 적용한 분산된 국내 문헌정보 DB의 연계 활용 모형 구축에 관한 연구

A study on constructing a database in liaison with the distributed bibliographic data using the Metadata Registry

박정훈, 이응봉 충남대학교 대학원 문헌정보학과

Park Jung-Hoon, Lee Eung-Bong

Graduate School of Library & Information Science Chungnam National University

현재, 다수의 조직에서 개별적으로 메타데이터 표준을 제시하고 구축하고 있다. 이로 인해 메타데이터 상호간 호환성의 문제가 발생하게 되고 각기 분산되어 구축된 메타데이터 이용 또한 원활하지 않게 되고 있다. 이를 해결하기 위해 본고에서는 메타데이터 레지스트리를 적용하여 국내 문헌정보 DB의 메타데이터 표준 유지와 통합 검색을 통한 연계 활용 모형을 구축하고자 한다.

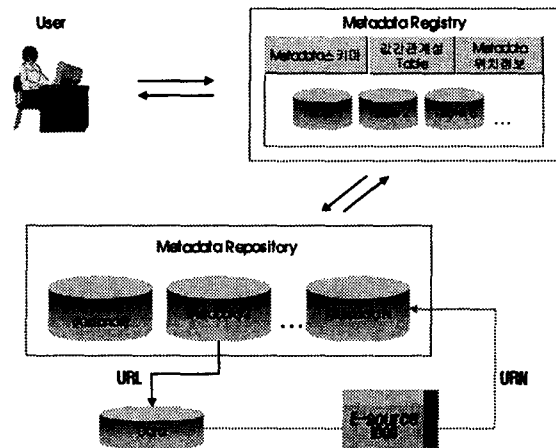
## 1 연구의 목적

메타데이터 레지스트리는 등록과 인증을 통하여 메타데이터 표준을 완성하고, 메타데이터를 공유하고자 하는 개념이다. 데이터 요소 레지스트리가 데이터 요소 수준에서 데이터의 공유를 가능하게 하는 것이라면, 메타데이터 레지스트리는 다양한 종류의 메타데이터 표준을 메타데이터 요소의 설명을 통하여 메타데이터의 공유를 가능하게 한다.

현재, 각기 개별적으로 필요에 의해 정의된 메타데이터 표준들은 상호 호환성의 문

제를 야기하고 있기 때문에 다수의 메타데이터 표준의 일관성을 유지시키고 각 메타데이터 표준에 존재하는 메타데이터 요소들의 의미를 명확히 하여 메타데이터의 호환성 및 상호운용성 향상을 위한 메타데이터 관련 연구는 필수적이라고 할 수 있다.

본 연구에서 다루고자 하는 내용은 분산되어 구축된 국내 문헌정보 DB의 연계 활용을 위하여 구현 운영중인 데이터 레지스트리들을 분석하고, 메타데이터의 표준 및 지속적인 관리를 위한 메타데이터 레지스트리 구축에 필요한 구체적인 구성 요소, 기능 및 관계를 정의하여 구축 모델을 제시하는 것이다.



〈그림 3-1〉 메타데이터 레지스트리 시스템 시스템 구성도

## 2 연구의 방법

일차적으로 국내 문헌정보 DB의 현황을 각종 문헌을 통하여 조사하고 ISO/IEC 11179에 대한 조사 및 표준화 추진현황을 조사한 후, 이들을 기반으로 문헌정보 데이터베이스에 대한 전반적인 현황을 분석한다. 이 과정에서 얻어진 결과를 통해 기존의 분산·구축된 메타데이터의 등록 및 관리를 위한 메타데이터 레지스트리 시스템의 연계 활용 모형을 구축한다. 또한 데이터 요소를 등록하고 검색할 수 있는 서비스 시스템들의 모형을 제시하여 전체 국내 문헌정보 DB의 연계 활용 모형을 구축한다.

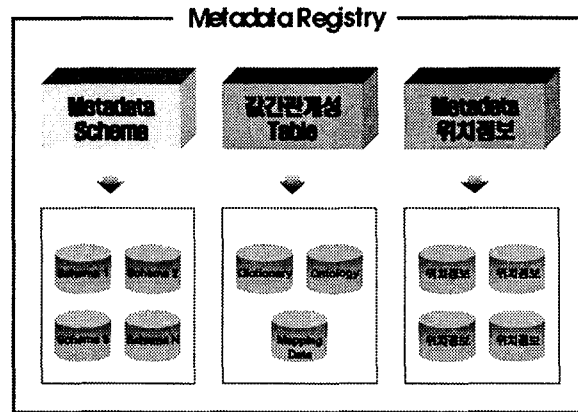
## 3 메타데이터 레지스트리

### 1) 개념

메타데이터 레지스트리란, 모든 종류의

데이터(즉, 데이터베이스, 메시지, 문서, 다른 응용 시스템에서 발견된 데이터)에 대한 의미, 구조, 교환 양식을 기록하는 정규 시스템이다. 또한 메타데이터 레지스트리는 기존에 정의된 규칙의 확장을 위한 규칙을 포함하도록 한다. 메타데이터의 다중 집합에 대한 레지스트리는 중복요소에 대한 정형화된 사상(mapping)이 포함될 수 있다. 다시 말하면, 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 및 관리하며 메타데이터의 명세와 의미의 공유를 목적으로 하는 것으로서 데이터의 표준화를 지원하고, 데이터의 공유를 가능하게 하는 것이다.

〈그림 3-1〉과 같이 메타데이터 레지스트리는 데이터에 대한 사용자와 관리자(user)들을 위하여 데이터를 공유하는데 필요한 기본적인 데이터 요소 특징들에 따라 메타데이터를 등록한다.



〈그림 3-2〉 메타데이터 레지스트리 내부 구성도

데이터 요소들에 대한 메타데이터는 메타데이터 레지스트리에 저장된다. 메타데이터 레지스트리는 데이터의 명확한 설명을 통하여 데이터의 표준화를 지원하고, 데이터의 공유를 가능하게 한다. 등록은 데이터 공유를 지원하기 위하여 메타데이터를 기록하는 작업이다. 이러한 기록 작업은 의미를 명확히 하기 위하여 데이터 요소단계에서 수행되어야만 한다. 메타데이터 레지스트리는 최종 사용자가 데이터 요소의 의미를 정확하고 모호하지 않게 이해할 수 있도록 한다.

또한 메타데이터 레지스트리는 데이터에 대한 사용자와 관리자들을 위하여 데이터를 공유하는데 필요한 기본적인 데이터 요소 특징들에 따라 메타데이터를 저장한다. 또한 식별자, 정의, 분류체계에 따라 메타데이터를 설명한다. 메타데이터 레지스트리는 사용자에게 데이터에 대한 정확한 의미를 발견할 수 있도록 한다. 또한, 사용자들이 데이터베이스로부터 데이터 값들을 검색하고자 한다면 원하는 데이터의 타입을

식별할 수 있도록 한다.

〈그림 3-2〉에 나타난 바와 같이 메타데이터 레지스트리는 크게 메타데이터 스키마, 값간 관계성 테이블, 메타데이터 위치 정보 등의 크게 세가지 부분으로 구성된다. 첫째, 메타데이터 스키마에는 등록된 메타데이터에 대한 각각의 스키마 정보를 가진 테이블들이 존재한다. 둘째, 값간 관계성 테이블에는 Dictionary, RDF를 이용하여 기술된 Ontology 정보, 각각 메타데이터들의 Mapping 정보가 담겨져 있다. 셋째, 메타데이터 위치정보에는 이용자들이 원하는 메타데이터를 검색했을 때 각각의 메타데이터 위치정보를 알려줌으로써 찾고자하는 메타데이터에 도달할 수 있도록 메타데이터 위치정보 데이터들이 저장되어 있다.

메타데이터 레지스트리의 주요 기능은 다음과 같이 정의할 수 있다. 첫째, 데이터의 요청과 등록을 유용하게 한다. 둘째, 데이터의 접근과 사용을 촉진한다. 셋째, 메타데이터에 의하여 기술된 특징을 이용하

여 데이터의 조작을 가능하게 함으로써 지능적 소프트웨어에 의한 데이터의 조작을 용이하게 한다. 넷째, CASE도구들과 Repository를 위한 데이터 표현 메타모델의 개발을 가능하게 한다. 다섯째, 전자 데이터 교환과 데이터 공유를 유용하게 한다.

또한, 메타데이터 레지스트리는 데이터에 대한 명확한 기술, 목록관리, 분석, 분류가 가능하도록 하고, 데이터 표준 개발 프로세스와 보급을 지원하고, 잘 정의된 데이터 요소와 값 영역의 원천을 제공하고, 데이터의 공유, 통합, 비교를 가능하게 하며, 메타데이터의 중앙 저장소 역할을 수행한다.

## 2) 관련표준

### ① 데이터 요소 표준 : ISO/IEC 11179)

이 국제 표준은 데이터를 이해할 수 있고 공유할 수 있도록 만들기 위한 표준화와 등록에 대한 내용을 설명하고 있다. 이 국제 표준에서는 데이터 요소의 표준화와 등록을 이용하여 기존의 데이터 관리 방법론들에 비하여 훨씬 적은 시간과 노력으로 공유 데이터 환경을 생성할 수 있는 방법을 제시한다. 즉, ISO/IEC 11179에서는 메타데이터 레지스트리에서 데이터의 의미, 구문, 표현을 표준화 할 수 있는 프레임워크를 제시하고 있다.

ISO/IEC 11179는 모두 6개의 부로 이루어져 있으며 각각의 내용은 다음과 같다.

Part 1 “데이터 요소의 명세와 표준화를 위한 프레임워크”에서는 데이터 요소들의

개념들에 대하여 소개하고 설명한다.

Part 2 “데이터 요소의 분류”는 데이터 요소 개념과 데이터 요소를 분류 방법에 의하여 분류하는 절차와 기술을 제시한다.

Part3 “기본적인 데이터 요소의 속성”에서는 데이터 요소들의 속성들에 대하여 정의한다. 여기에서는 데이터 요소들의 사용과는 무관한 기본 속성들만을 언급하고 있다. 데이터 처리와 데이터 상호 교환에 있어서 정확성, 신뢰성, 상호운용성은 사용자와 제공자에 있어서 점차로 중요해지고 있고, 이러한 것을 지원하기 위해서는 교환되는 데이터가 데이터 요소에 대하여 공통된 개념을 가지고 있어야 한다.

Part4 “데이터 정의의 공식화를 위한 규칙과 안내”에서는 모호하지 않은 데이터 요소를 정의하는 지침을 제시한다. 데이터 요소의 공유를 위하여 잘 정의되어진 정의는 필수적인 요소가 된다. 만약 모든 데이터 사용자들이 정확하고 잘 정의되어진 데이터 정의를 통하여 통신을 한다면 많은 문제점들이 사라지게 될 것이다.

Part5 “데이터 요소를 위한 명명과 식별의 원칙”은 데이터 요소의 식별에 대한 지침이다. 식별은 특정 데이터 요소에 대한 지시자 부여나 식별하는 작업을 말한다. 식별은 숫자적인 식별자의 부여를 포함한다.

Part6 “데이터 요소의 등록”은 중앙등록 위원회를 통하여 데이터 요소를 등록하는 방법과 유일한 데이터 요소에 대한 식별자 부여 방법에 대한 사항이다. 등록된 데이터 요소의 유일성은 등록 위원회 식별자, 등록

위원회에서 부여한 데이터 요소의 식별자, 그리고 버전의 조합에 의하여 결정되어진다. 각각의 레지스트리는 등록 위원회에 의하여 관리된다. 레지스트리는 만약 필요한 데이터 요소가 그안에 존재한다면, 쉽게 접근할 수 있도록 색인되어 구성하여야 한다.

#### ② 메타모델 : X3.285(ANSI)

이 표준은 데이터 레지스트리의 구조를 명시한다. 구조는 함축적인 데이터 모델의 형태로서 설명된다.

이 표준에 표현된 데이터 레지스트리의 구조는 관리 책임 부분, 명명과 식별, 분류, 데이터 요소 개념 관리, 개념 영역과 값 영역의 관리, 데이터 요소의 관리 등의 6개의 부분으로 구성된다.

관리 책임 부분은 데이터 요소이거나 정의와 재사용이나 공유를 위한 명세를 요구하는 구성 요소이며, 조직 사이에서 사용을 위한 관리를 요구하는 중요한 구성 요소이다. 관리 구성 요소는 데이터 요소, 분류 스키마, 값 영역, 규칙이나 분류된 구성 요소를 위한 일반화이다.

명명과 식별은 관리되는 구성 요소의 이름과 그 이름의 세력 범위를 제공하는 이름 문맥을 관리하기 위해 사용되는데, 고유 식별자의 부여와 데이터 요소의 이름과 관련되어 있다.

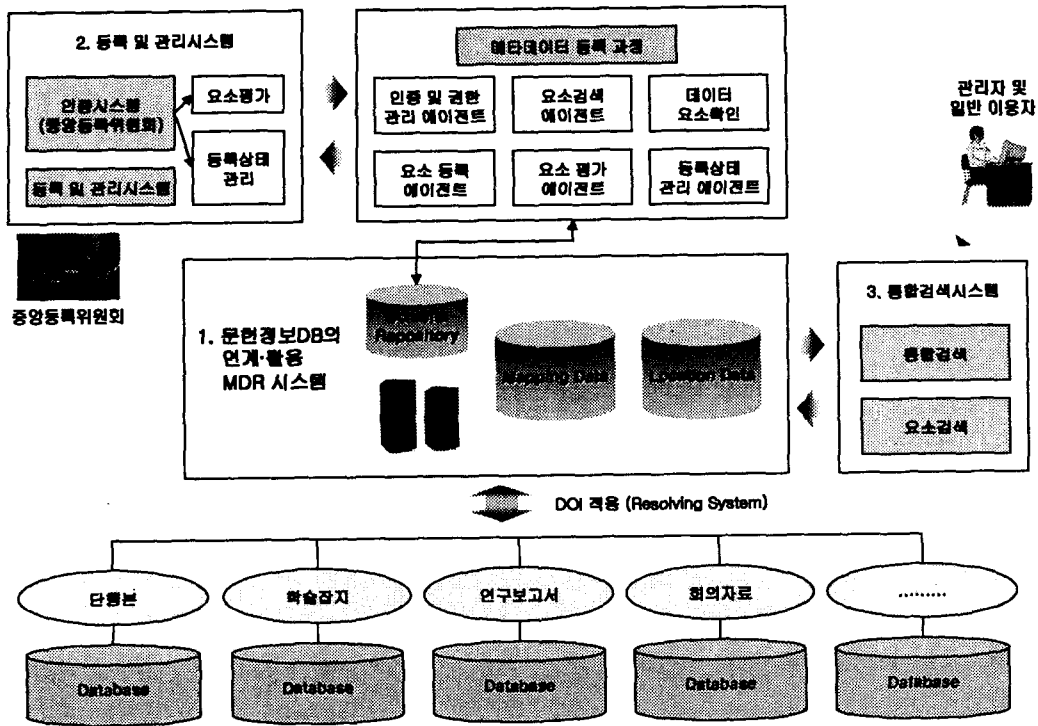
분류는 분류 스키마와 분류 스키마 내에 존재하는 저장소 구성 요소를 관리하기 위하여 사용되는데, 등록 가능 요소들의 분류 체계 관리 등과 관련된다.

데이터 요소 개념 관리는 데이터 요소의 의미에 해당하는 개념 영역에 대한 관리이다. 데이터 요소에 대한 관리는 데이터 요소 자체에 대한 관리이고, 값 영역과 개념 영역의 관리는 실질적인 표현에 해당하는 부분과 관련된다.

## 4 MDR을 적용한 분산된 문헌정보 DB 연계 활용 시스템의 구축 방안

효율적인 정보 공유가 이루어지기 위해서는 정보의 의미와 표현방법 등에 대한 일정한 약속이 선행되어야 한다. 메타데이터 레지스트리를 이용하여 각 문헌정보 DB의 메타데이터를 등록하고 그 표준을 유지함으로써 각각 상이한 형태로 구축되어 있는 DB를 항목별 상세검색이 가능한 지능적인 통합검색 및 연계시스템을 구축할 수 있다.

이것은 분산되어 구축된 각 문헌정보 DB를 대상으로 하여 통합된 데이터 모델을 설계하고 ISO/IEC 11179에 의하여 각 데이터요소를 정의하여 대상 도메인의 데이터베이스를 통합하고, 장기적으로는 모든 지식정보의 통합관리 및 검색이 가능하게 된다. 학술지, 연구보고서, 회의자료, 단행본 등에 대한 데이터 요소를 명확히 정의하고 데이터 스키마를 메타데이터 레지스트리에 등록하여 메타데이터의 표준안을 유지·관리함으로써 중복투자 및 자원낭비의 비효율을 제거할 수 있다.



<4-1> 국내 문헌정보 DB의 연계활용 모형

그리고, 향후 구축되는 문헌정보 DB에 대한 검증과 각 요소들에 대한 의미 설명을 제공해주는 국가적인 메타데이터 관리 시스템을 구현할 수 있다.

<그림 4-1>에서의 국내 문헌정보 DB의 연계 활용 모형은 각 문헌정보 DB를 대상으로 하여 통합된 데이터 모델을 설계하고 ISO/IEC 11179에 의하여 각 데이터요소를 정의하여 등록하는 메타데이터 레지스트리 시스템을 기반으로 각 문헌정보 DB의 메타데이터들을 등록할 때 중앙등록위원회의에서 메타데이터의 검증과 요소평가, 등록된 메타데이터의 관리를 수행하는 등록 및 관리시스템, 그리고 이용자가 실제 데이터

베이스를 검색 및 이용할 수 있는 통합검색시스템으로 구성된다.

문헌정보 DB의 MDR 시스템은 데이터 베이스 스키마를 등록해 놓은 스키마 Repository와 값간 관계성 Table 등을 가지고 있는 Mapping Data, 그리고 데이터의 위치정보를 가지고 있는 Location Data로 구성된다.

등록 및 관리시스템에서는 중앙등록위원회에서 요소를 평가하여 인증시스템에 의해서 등록을 하고, 등록된 데이터를 관리시스템에서 관리하면서 메타데이터의 표준화를 유지·관리하게 된다.

메타데이터의 등록 과정은 인증 및 권한

을 확인한 후 요소검색을 실시하고 데이터 요소 및 등록상태를 확인하고 등록하고자 하는 요소를 평가해서 등록하게 된다.

통합검색시스템에서는 MDR 시스템에 있는 스키마 정보와 Location Data를 이용하여 질의 분석, 스키마 분석, 질의 확장, 질의 분배를 거쳐 데이터베이스에 접근하고 그 결과를 처리 및 통합하여 이용자에게 제공하는 역할을 수행한다. 이용자들에게 제공하는 원문은 DOI (Digital Object Identifier)와 Resolving System을 이용하여 이용자에게 제공하게 된다.

DOI는 디지털 콘텐츠(원문) 접근 및 검색 효율성을 향상시키고 저작권의 보호 및 이용의 편의성을 제고하는 것으로서 각 원문에 DOI를 부여한 후 DOI Resolving system을 통하여 이용자에게 제공하게 된다.

본 연구에서는 각 시스템 구성요소들의 기능을 명세하여 각 시스템별 상세 구조를 설계한다. 설계한 시스템을 기반으로 메타데이터의 검증 및 등록과정부터 이용자가 검색한 후 그 결과를 다시 이용자에게 제공하게 되는 원문이용까지의 전 과정을 도식화하여 국내 문헌정보 DB의 연계활용 모형을 제시한다.

## 5 결 론

분산된 국내 문헌정보 DB의 연계 활용을 위하여 그동안 Z39.50 등의 연구가 진행되어 왔지만 시스템의 복잡성으로 인한 사용의 어려움 때문에 일반적으로 적용하

기 힘든 한계가 있는 등 여러 가지 문제가 발생하고 있다.

그러나 메타데이터를 유지하고 관리하며, 메타데이터의 표준화를 지원하고, 데이터의 공유를 가능하게 하는 메타데이터 레지스트리 시스템을 이용함으로써 지금까지 이루어지지 못했던 문헌정보 DB의 메타데이터 표준화를 달성함과 동시에 분산되어 구축되어오던 문헌정보 DB의 메타데이터를 메타데이터 레지스트리 시스템에 등록시키고 통합검색시스템과 DOI Resolving System을 적용시킴으로써 원문 제공뿐만 아니라 각각의 문헌정보 DB를 연계 활용할 수 있는 시스템이 가능하다고 할 수 있다.

## 6 참고문헌

- [1] “과학기술 지식정보 공유 및 표준화 관리 모델 개발연구”, 고려대학교, 2002. 11
- [2] “메타데이터 레지스트리 구축을 위한 프레임워크 표준 연구”, 한국전산원, 2000. 12
- [3] “공공부문의 웹 기술 표준화 방안”, 한국전산원, 2002. 12
- [4] ISO/IEC 11179 Part 1-6 Metadata Registries, 1998
- [5] ANSI X 3.285 : Metamodel for the Management of Shareable Data, 1998
- [6] <http://www.epa.gov/edr/>
- [7] <http://www.dstc.edu.au/Research/Projects/rdf/>