

메타데이터 표준화와 메타데이터 레지스트리

고 영 만 (성균관대학교 문헌정보학과 교수)

목 차

- | | |
|-----------------------------|--------------------------|
| 1. 메타데이터와 메타데이터 표준화 | 3. 메타데이터 레지스트리 표준화 |
| 1.1 메타데이터 | 3.1 메타데이터 레지스트리 국제 표준 |
| 1.2 메타데이터 표준화 | 3.2 메타데이터 레지스트리 표준 적용 사례 |
| 2. 메타데이터 상호운용성과 메타데이터 레지스트리 | 4. 국가 메타데이터 레지스트리 (NMR) |
| 2.1 메타데이터 상호운용성 | 4.1 국가 메타데이터 레지스트리의 개념 |
| 2.2 메타데이터 레지스트리 | 4.2 국가 메타데이터 레지스트리의 추진체계 |
| | 4.3 국가 메타데이터 레지스트리의 기대효과 |

1. 메타데이터와 메타데이터 표준화

1.1 메타데이터

메타데이터는 정보자원의 속성들을 기술하는 데이터로서 ‘데이터에 대한 데이터’로 간략하게 정의된다. 메타데이터의 속성으로는 내용에 관한 사항, 다른 정보자원과의 관계, 정보자원의 지적 특성에 관한 사항, 정보자원의 물리적 형식 및 생성 일자 등 정보자원을 식별할 수 있는 사항을 들 수 있다. 메타데이터는 이러한 속성 정보를 구조화하고 이 구조를 통해 정보자원의 특성을 묘사한다.

메타데이터의 기능이 정보자원을 인식하고 구조화하는 것이므로 메타데이터의 범위를 편목 기입과 같은 전통적인 서지데이터까지 확장해서 설명하는 경우도 있다. 정보를 인식하고 구조화한다는 측면에서 부분적으로 사실이다. 그렇지만 메타데이터의 구조화 대상은 “네트워크 상의 전자정보원”으로 한정하는 것이 일반적이다. 이런 의미에서 메타데이터가 서지데이터와 다른 점은 다양한 접근점과 위치정보를 통해 적절한 응용소프트웨어와 정보통신망의 도움으로 정보자원 자체에 직접 접근할 수 있게 해주는 점이라 할 수 있다.

정보자원에 대한 메타데이터를 생성시킬 경우 정보자원을 보다 정확하게 검색할 수 있어서 전문을 대상으로 하는 키워드 방식에서 나타나는 방대한 검색 결과의 문제점을 극복할 수 있으며, 텍스트 자료뿐만 아니라 이미지, 동영상 자료들의 검색도 가능해진다. 또한 태그(Tag) 부여와 같은 정보 가공 작업을 통해 부가 가치를 생성시킬 수 있으며, 정보 검색 외에도 메타데이터가 언제 생성되었고 갱신되었는지, 책임자가 누구인지 등과

같은 관리용 메타데이터 요소를 가질 수 있으므로 메타데이터 자체의 관리를 체계적이고 효율적으로 할 수 있다.

1.2 메타데이터 표준화

메타데이터가 지니는 검색과 관리의 효율성에 따라 여러 MARC, DC, MODS, ONIX 등의 메타데이터 포맷이 도서관 분야에서 적용되고 있으며, 다른 분야에서도 CERIF, INDEX, GDAS와 같은 다양한 형식의 메타데이터 포맷이 개발되어 이를 토대로 하는 시스템이 구축되고 있다. 그러나 서로 상이한 환경과 요구에 따라 개발된 메타데이터 구조와 형식은 오늘날 메타데이터 간의 상호운영성을 저해하는 요소로 작용한다. 서로 다른 형식의 메타데이터를 사용할 경우 데이터 요소와 표현의 불일치에 따라 동일한 내용의 정보자원이 상이한 형식으로 표현되며 다른 형식의 메타데이터에는 접근할 수 없기 때문이다. 이는 동일한 도메인(주제분야)에서만 발생하는 것이 아니라 서로 다른 도메인 간에도 동일하게 적용되어 데이터의 중복, 오류 등의 비효율성을 누적시키는 결과로 이어진다. 메타데이터 표준화는 이러한 문제점을 극복하고 시스템 간 정보자원의 상호운영성을 확보하기 위한 것으로서, 표준화가 이루어질 경우 메타데이터와 해당 정보자원 관리 및 검색 측면에서의 효율성이 증가된다.

2. 메타데이터 상호운영성과 메타데이터 레지스트리

2.1 메타데이터 상호운영성

메타데이터 상호운영성은 이용자 측면에서 볼 때 다양한 출처에서 생성되는 정보자원을 성공적으로 검색하고 그 결과를 확신할 수 있게 하는 것이라 할 수 있다. 시스템 관점에서는 두 개 이상의 시스템이 정보를 교환하고 교환된 정보를 특별한 노력 없이 이용할 수 있게 하는 기능, 서로 다른 메타데이터를 사용하는 시스템들이 각각의 메타데이터를 이해할 수 있는 기능, 또는 메타데이터 스키마에 상이한 어휘로 규정된 기술 요소들의 의미를 이해할 수 있는 기능이라 할 수 있다. 메타데이터의 상호운영을 확보하기 위한 방법으로는 크게 자원을 하나의 표준적인 메타데이터로 통합하여 표현하는 방법, 자원의 특성을 감안하여 다양한 메타데이터 형식과 기술 구조를 인정하고 상호 매핑을 통해 해결하는 방법, 그리고 메타데이터 온톨로지 즉 메타데이터 레지스트리(Metadata Registry)에 의한 해결 방법의 세 가지를 들 수 있다.

하나의 메타데이터로 통합시키는 방법은 모든 형태의 네트워크 자원을 하나의 표준적인 메타데이터로 표현하는 것이 가장 이상적이라는 관점에서 출발한 것이다. 다양한 정보

자원과 메타데이터를 하나의 틀 안에서 통합하여 조직할 수 있는 가장 효율적인 메타데이터 포맷을 선정하고, 이를 표준적인 통합 메타데이터 포맷으로 사용하자는 것으로 MARC(Machine Readable Cataloging)와 더블린코어(Dublin Core)가 대표적인 예이다.

메타데이터의 매핑은 정보자원의 특성을 감안하여 다양한 메타데이터 형식과 기술 구조를 인정해야 한다는 관점에서 메타데이터 상호운영의 문제점을 해결하고자 하는 방식이다. 오늘날 각기 다른 기술 구조를 가진 다양한 메타데이터를 상호운영성의 입장에서 통합하기 위한 연구가 이루어지고 있으며, 상호참조테이블 방식과 범용 메타데이터 통합 방식이 있다. 상호참조테이블 방식에는 USMARC과 DC, TEI, EAD의 매핑 테이블, DC와 다른 메타데이터 형식과의 참조 테이블, 공통 메타데이터 기술 집합과의 매핑 등이 있다. 그러나 의미상의 일치를 위해서는 전문가의 지적 능력이 필요하고, 메타데이터의 잦은 수정에 의한 유지관리의 부담, 정확한 1:1 대응이 이루어지지 않는 문제점을 지니고 있다. 범용 메타데이터 통합방식으로는 W3C에서 개발한 개념적 차원의 통합 구조 RDF가 실제에서 구현되고 있다.

2.2 메타데이터 레지스트리

메타데이터 레지스트리 역시 메타데이터의 상호운영성을 확보하기 위한 목적으로 고안된 것이다. 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 관리하며, 메타데이터의 명세와 의미의 공유 즉 ‘공유되는 개념의 정형적 명시적 명세화’를 통해 메타데이터 셋 또는 메타데이터 요소 간의 호환성을 유지시킨다. 구체적으로 표현하면 서로 다른 데이터베이스가 같은 개념에 대해 서로 다른 식별자(태그) 혹은 서로 다른 단어를 사용할 경우 이를 해결해 주기 위해 공유되는 개념화를 정형적, 명시적으로 명세화한 결과물인 온톨로지가 집합되어 있는 메타데이터 관리시스템이라 할 수 있다. 예를 들어 표제와 제목이라는 상이한 데이터 요소명칭을 사용하는 두 개의 데이터베이스에 들어있는 정보를 비교하거나 통합하려는 프로그램에서는 표제와 제목이 같은 의미를 지칭하는 메타데이터 요소/식별자라는 것을 알아야 하며, 메타데이터 레지스트리에는 바로 이러한 관계가 정형적, 명시적으로 명세화 된다.

메타데이터 레지스트리에서는 규격화된 방법론에 의해 메타데이터 요소가 등록, 승인, 삭제된다. 이러한 의미에서 메타데이터 레지스트리는 사전 예방에 의한 문제해결을 강조하지만 기존의 메타데이터 요소 간의 상호운영성을 확보한다는 측면에서는 사후 관리를 지향하는 방식이라 할 수 있다. 그러나 상이한 프레임워크(관리체계, 관리방식, 관리내용 등)를 가진 메타데이터 레지스트리가 사용될 경우 메타데이터의 상호 운영에 새로운 문제가 발생하게 된다.

3. 메타데이터 레지스트리 표준화

메타데이터 레지스트리는 데이터의 사용자와 관리자들이 데이터를 공유하는 데 필요한 기본적인 메타데이터를 저장하며 메타데이터의 식별자, 정의, 분류 체계에 따라 각 메타데이터의 이력을 설명한다. 메타데이터 레지스트리는 등록된 데이터 요소의 의미와 이력, 관련성 등이 정확하고 모호하지 않게 이해될 수 있도록 관리되었을 때 그 기능이 보장될 수 있다. 따라서 데이터의 요청과 등록 체계를 규격화하여 메타데이터 레지스트리를 정형적으로 관리하기 위한 표준화된 방법론의 필요성이 대두되었으며, 오늘날 국제표준(ISO/IEC 11179 - Metadata Registry)으로 제정되어 수정과 보완작업이 진행되고 있다.

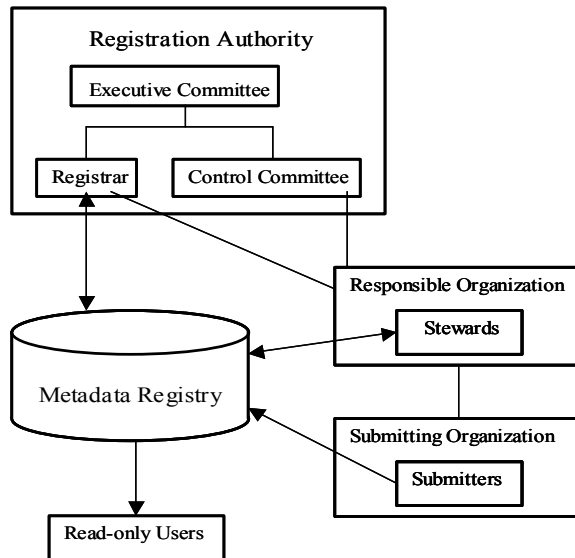
3.1 메타데이터 레지스트리 국제표준 : ISO/IEC 11179 - Metadata Registry

ISO/IEC 11179는 ISO/IEC JTC1 SC32의 WG2에서 개발한 메타데이터 레지스트리에 관한 국제 표준이다. ISO/IEC 11179는 메타데이터 레지스트리 구축에 관한 구체적인 시스템 아키텍처와 데이터베이스 설계 등을 명시하고 있지는 않으나, 기술된 개념 및 논리적 데이터 모델에 따라 메타데이터 레지스트리를 구축할 수 있게 하며 총 여섯 개의 세부 규격으로 나뉘어져 있다 :

- 제1부 데이터 요소의 명세와 표준화를 위한 프레임워크 : 데이터 요소의 개념
- 제2부 데이터 요소의 분류 : 데이터 요소 개념과 데이터 요소를 분류방법에 의하여 분류하는 절차
- 제3부 MDR 메타모델 : 메타데이터 레지스트리의 메타모델
- 제4부 데이터 정의의 공식화를 위한 규칙과 안내 : 명확하게 데이터 요소를 정의하는 지침
- 제5부 데이터요소를 위한 명명과 식별의 원칙 : 데이터 요소의 식별에 대한 지침
- 제6부 데이터 요소의 등록 : 중앙등록위원회를 통하여 데이터 요소를 등록하는 방법과 유일한 데이터 요소에 대한 식별자 부여 방법

ISO/IEC 11179는 데이터 요소(메타데이터)의 생성, 등록, 관리를 지원함으로써 시스템 또는 조직 간의 정보 공유를 지원하며, 데이터의 사용자들이 데이터의 의미, 표현, 식별 등에 대하여 쉽게 이해할 수 있도록 한다. 또한 메타데이터 수집의 일관적 모델 제공하고 데이터에 대한 명확한 기술, 목록관리, 분석, 분류가 가능하도록 하며, 데이터 표준 개발 프로세스와 보급을 지원하고 잘 정의 된 데이터 요소와 값 영역의 원천을 제공함으로써 데이터의 공유, 통합, 비교를 가능하게 한다. 특히 등록관리기관(Registration Authority : 제출된 데이터 요소에 등록번호를 부여하고 등록된 데이터 요소를 적절히 관리하는 조직), 등록기관(Submitting Organization : 데이터를 RA에 등록하고 등록된 데이터의 기본 속성과 추가적인 정보를 제공하는 조직), 자문기관(Responsible Organization : RA의 요

청에 따라 제출된 데이터 요소의 문맥, 이름, 허용가능 값 등에 대한 자문을 제공하는 조직)을 통해 메타데이터의 상태와 활동주기를 관리하도록 되어 있다 (그림 1 참조).



<그림 1> IISO/IEC 11179 MDR의 운영 개념도

ISO/IEC 11179에서의 데이터 상태는 등록, 인증, 표준의 세 단계로 관리되며 데이터 활동주기는 불완전, 등록, 인증, 표준, 퇴역의 다섯 단계로 나뉘어 관리된다.

- 데이터 상태 :
 - 등록된 데이터 요소 (Recorded Data Element)
 - 인증된 데이터 요소 (Certified Data Element)
 - 표준화된 데이터 요소 (Standardized Data Element)
- 데이터 활동주기 :
 - 불완전 단계 (Incomplete : 데이터 요소가 필수 속성을 가지고 있지 않는 단계)
 - 등록 단계 (Recorded : 데이터 요소의 내용이 ISO 11179에서 제시하는 품질 요구사항에 못 미치는 단계)
 - 인증 단계 (Certified : 데이터 요소의 내용이 ISO 11179에서 제시하는 품질 요구사항에 부합된 단계)
 - 표준 단계 (Standardized : 등록기관에 의하여 인증된 데이터 요소가 만들어진 단계)
 - 퇴역 단계 (Retired : 데이터 요소가 더 이상 사용되지 않는 단계)

2.2 메타데이터 레지스트리 표준 적용 사례

메타데이터 레지스트리 표준을 적용하여 시스템을 구축한 대표적 사례로는 해외의 경우 미국의 USHIK, ITS, EDR을 들 수 있으며, 국내에서는 IMR을 들 수 있다.

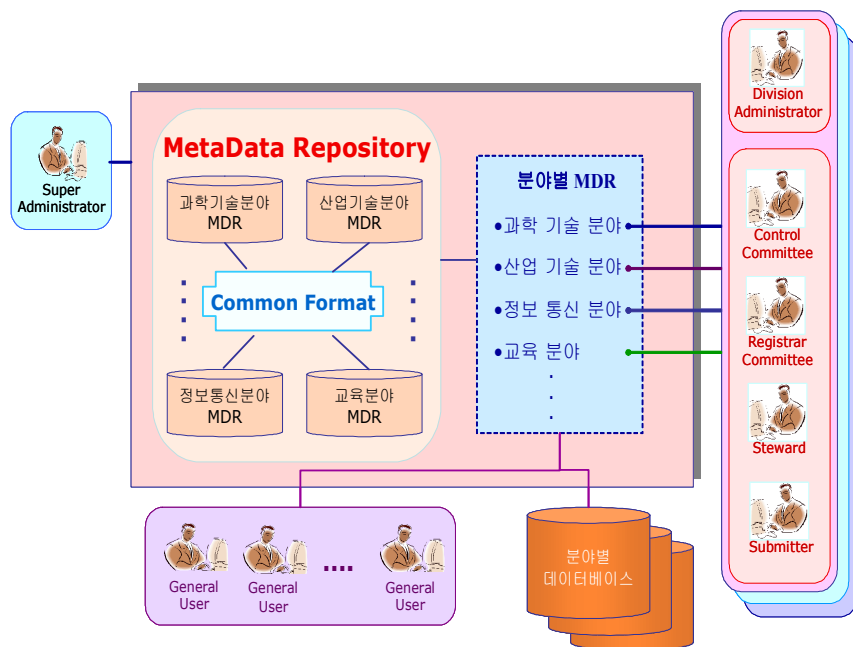
(1) 해외 사례

- USHIK (US Health Information Knowledgebase, <http://www.ushik.org/registry/x/registry.html>) : USHIK은 미국 국방성(DoD)의 건강국(Health Affair : HA)과 의료지원서비스센터(Centers for Medicare and Medicaid Services : CMS)가 공동으로 수행한 프로젝트에서 구축된 메타데이터 레지스트리 시스템이다. 건강 표준개발기관(Healthcare Standard Development Organization : SDO)들과 관련 기관들 사이의 데이터 요소를 분류하고 조화하는 작업을 하였으며, 'Health Insurance Portability and Accountability Act (HIPAA)'의 데이터 요소를 사용하여 메타데이터 레지스트리를 구축하였다. America Standards Committee (X12), Health Level Seven (HL7), Health Industry Business Communications Council (HIBCC), National Council for Prescription Drug Program (NCPDP), Medical Device Communications Industry Group (IEEE 1073), American Dental Association (ADA) 등의 기관이 참여하고 있다.
- ITS (Intelligent Transportation System, http://www.standards.its.dot.gov.net/data_reg.htm) : ITS의 데이터 레지스트리(ITS DR)는 미국 교통부(U.S. Department of Transportation)에서 수행한 지능형 교통시스템인 ITS 내 각 시스템간의 상호운영성을 보장 위해 개발된 것으로 ISO/IEC 11179 MDR 표준을 토대로 구축하였다. 50개 주와 지역 교통국이 ITS 시스템 내부에서 혹은 중앙정부와 ITS 외부 시스템이 가지고 있는 데이터를 의미에 따라 교환하고 상호작용한다. American Association of State Highway and Transportation Officials (AASHTO), American National Standards Institute (ANSI), American Society for Testing & Materials (ASTM), Consumer Electronics Association (CEA), Institute of Electrical and Electronics Engineers (IEEE), Institute of Transportation Engineers (ITE), Society of Automotive Engineers (SAE) 등의 기관이 참여하고 있다.
- EDR (Environmental Protection Agency Metada Registry, <http://www.epa.gov/edr/>) : 미국환경청(U.S. Environmental Protection Agency : EPA)에서 ISO/IEC 11179 표준을 토대로 구축한 메타데이터 레지스트리이다. EDR은 환경 관련 기관의 표준들을 기록할 수 있게 하고 기관과 사용자 간의 데이터 공유를 가능하게 하기 위하여 이해가 쉽고 신뢰성이 높은 환경관련 자료를 제공하고 있으며, 다음과 같은 기능을 가진 자세한 설명을 통하여 환경관련 데이터의 의미 활용도를 높이고 있다 :
 - Data Dictionary : EDR은 데이터 이름, 정의, 형식과 관계들에 관한 정보 제공
 - Directory : 데이터에 대한 책임을 가지고 있는 담당 기관에 대한 정보 제공

- Reference Book : 데이터에 대한 세부 설명 정보 제공
- Road Map : 어떤 데이터를 포함하는 데이터베이스를 어떤 응용이 사용되는가와 같은 정보의 관계와 위치 정보 제공

(2) 국내 사례

- IMR (Industry Metadata Registry, http://isv.kisti.re.kr/new/new_index.html) :
IMR은 한국과학기술정보연구원(KISTI)에서 운영하고 있는 산업기술정보 메타데이터 레지스트리로서 산업자원부 기술표준원의 지원으로 산업기술정보 분야의 메타데이터를 표준화하여 보급하기 위한 목적으로 구축되었다. IMR은 국제표준 ISO/IEC 11179에 의거 데이터의 의미, 구문, 표현의 표준화 프레임워크 제시하고 있으며, 메타데이터의 명세와 의미를 공유하기 위해 메타데이터의 등록과 인증을 통한 표준화된 메타데이터를 유지, 관리함으로써 산업기술정보 분야에서 필요한 데이터의 호환성을 유지하고 있다 (그림 2 참조).



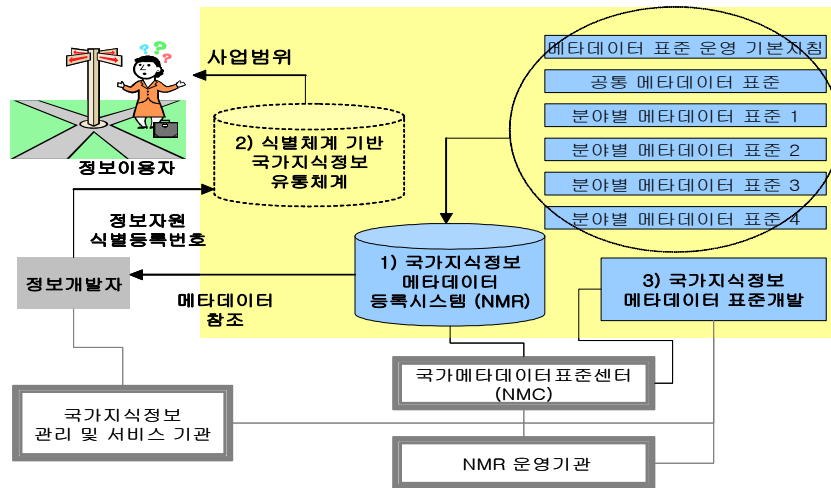
<그림 2> KISTI의 IMR 운영 개념도

국내에서는 IMR 외에 한국정보문화진흥원(KADO)에서 종합정보센터에 구축된 국가지식정보 메타데이터 관리 표준 모델 개발 사업을 추진하고 있다.

4. 국가 메타데이터 레지스트리(National Metadata Registry : NMR)

4.1 국가 메타데이터 레지스트리의 개념

지식정보의 데이터베이스의 구축에 있어서 표준화된 정형적 모형을 따르지 않고 제작자와 제작기관에 따라 상이한 형식으로 개발되어 데이터가 유통될 경우 동일 분야 데이터베이스들 간에도 지식정보의 공유가 어렵게 되고 물적, 인적 자원의 낭비가 많아진다. 국가 메타데이터 레지스트리(NMR)는 이러한 손실을 방지하기 위해 국내에서 생성되는 모든 분야의 지식정보들에 대한 메타데이터 표준을 개발하고 관리할 수 있도록 국가적 차원의 조직 설립을 통해 메타데이터의 정확성과 신뢰성을 유지하고 이를 주도적으로 관리하는 시스템이라 할 수 있다 (그림 3 참조).



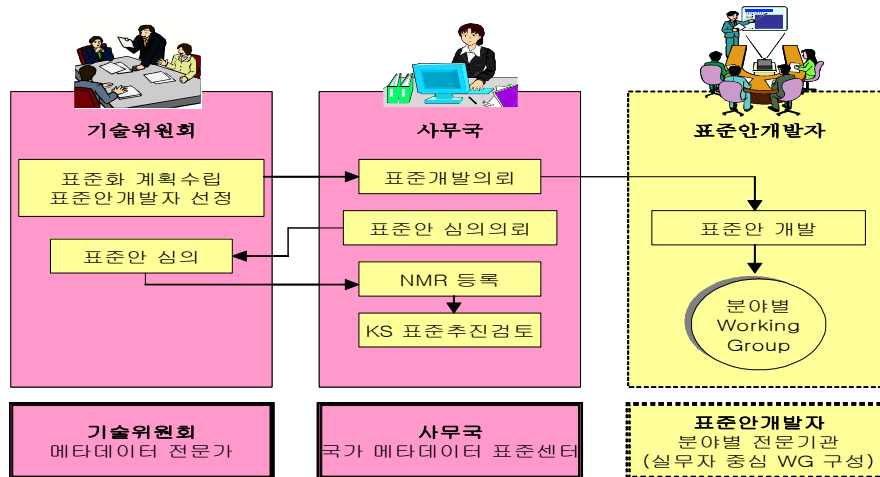
<그림 3> NMR 개념도

4.2 국가 메타데이터 레지스트리(NMR)의 조직 체계

핵심 조직으로 국가메타데이터표준센터(NMC : National Metadata Center)를 설립하여 국내 지식정보의 생성, 관리, 유통 및 이용에 필요한 데이터의 표준을 개발하고, 포괄적이고 권위있는 메타데이터 레지스트리 시스템을 관리하도록 하는 것이 필요하다. NMC는 NMR을 구축하여 운영하며, 각 기관들 간에 지식정보의 효과적인 공유를 촉진하기 위해 데이터 표준의 제정과 실행을 지원하고 지식정보자원의 공동 이용 체계와 제공되는 서비스의 일관성을 유지하는 역할을 한다. 기술위원회는 메타데이터 분야 전문가로 구성하며 표준화 계획을 수립하고 개발된 표준안을 심의하는 역할을 한다 (그림 4 참조). 이를 위해서 제도적 근거 규정의 마련되어야 하며 국제 및 국내 메타데이터 표준화 조직과의 긴

밀한 연계를 통한 표준화 추진이 요구된다 :

- ISO/JTC1 SC32 (Data Management and Interchange) WG2 : 메타데이터 표준
- TTA(한국정보통신기술협회) PG406 : 메타데이터 프로젝트 그룹
- 메타데이터 표준화 포럼



<그림 4> NMR 운영을 위한 조직체계

국가 메타데이터 표준센터(NMC)의 구체적 업무로는 표준화된 데이터 요소와 데이터 값셋의 제공, 데이터의 중복 확인, 효과적인 정보 수집, 프로그램 간의 정보의 일관성 및 공유 등을 들 수 있으며, 수행해야할 세부 사업은 다음과 같다 :

- 국가메타데이터등록시스템(NMR : National Metadata Registry) 구축
- 식별체계(DOI) 기반 산업기술정보 유통체계 구축
- 국가지식정보 메타데이터 표준 개발 의뢰

4.3 국가 메타데이터 레지스트리의 기대효과

국가 메타데이터 레지스트리는 국가지식정보의 메타데이터 표준화 기반을 주도하는 시스템으로서 시스템이 구축될 경우 다음과 같은 효과가 기대될 수 있다 :

- 국가지식정보자원의 유통의 상호운영성 확보 : 국내에서 구축되는 공공부문 지식정보 유통의 중심적 기능을 수행함으로써 정보자원 유통의 상호운영성이 확보되고, 국내에서 발생하는 모든 지식정보 데이터들에 대한 요소의 표준화가 가능해진다.
- 데이터베이스의 정확성과 신뢰성 향상 : 표준화된 데이터 요소를 기반으로 하는 데이터베이스 구축을 통해 불일치성을 사전에 제거하고 통일성을 제공함으로써 국내 유사 데이터베이스들 간의 일관성 유지 및 관련 기관 간의 데이터 교환과 공유 기능이 향상 된다.

- 통합 검색 시스템을 활용한 검색 효율의 극대화 : 국내에서 분산적으로 제공되는 이질적인 검색 인터페이스 통합에 활용할 수 있으며, 이용자의 접근을 보다 용이하고 일관성 있게 하는 인터페이스를 제공할 수 있게 됨으로서 검색 효율과 서비스의 질이 극대화된다.

참고문헌

- 고영만. 2005. 메타데이터와 온톨로지. 제3회 메타데이터 표준화 세미나. 산업자원부 기술 표준원. 한국기술센터, 2005.06.03.
- 서태설. 2004. 디지털도서관과 메타데이터 표준화. 제 7회 디지털도서관 컨퍼런스. 한국데이터베이스진흥센터. COEX, 2004.11.19
- 송인석. 2004. 산업기술정보 메타데이터 표준. 제2회 메타데이터 표준화 세미나. 산업자원부 기술표준원. 한국과학기술회관, 2004.12.10
- 옥철영, 2004. 한국어 정보처리와 온톨로지. 한국어정보처리연구회 2004 동계 튜토리얼. 성신여대, 2004.02.13
- 한국과학기술정보연구원. 2004. 메타데이터 표준화 지침.
<http://203.250.194.162/new/new_index.html>
- 한국전산원. 2004. 지식정보자원관리 메타데이터 구축 지침.