

ISO/IEC 11179에 따른 산업기술정보 메타데이터 표준화*

Standardization of Industrial Information Metadata Based on ISO/IEC 11179

남 영 광** · 서 태 설*** · 황 상 원****

Young-Kwang Nam · Tae-Sul Seo · Sang-Won Hwang

차 례

1. 서 론	5. 산업기술정보 메타데이터 레지스트리 (IMR) 개발결과
2. 관련 연구	6. 개발결과 고찰
3. ISO/IEC 11179의 개요	7. 결 론
4. 산업기술정보 메타데이터 표준안 개발	· 참고문헌

초 록

본 논문에서는 ISO/IEC 11179에 따라 산업기술정보 메타데이터 표준안을 작성하고, 메타데이터를 하나의 시스템에서 온라인으로 표준화할 수 있는 산업기술정보 메타데이터 레지스트리(IMR)를 개발하였다. 메타데이터 표준안 개발의 경우, 국내에서는 처음으로 ISO/IEC 11179 표준에 따라서 데이터 요소를 정의하여, 4개의 메타데이터 표준안을 개발하였다. 산업기술정보 메타데이터 레지스트리는 4개의 분야별 MDR로 구성되도록 구축하였다. 그에 따라 통합관리자 모듈과 분야별 관리자 모듈이 함께 개발되었다. 메타데이터 표준화 조직이 시스템을 통해 데이터 요소의 등록과 투표과정을 수행하여 새로운 데이터 요소를 메타데이터 레지스트리에 제공할 수 있다. 이용자는 산업기술정보 메타데이터 레지스트리를 분야별로 검색하여, 자신들의 데이터베이스를 구성할 수 있으며, 다양한 데이터베이스 스크립트를 내려 받을 수 있다. 시스템 개발은 ORACLE 데이터베이스를 기반으로 Java JSP를 이용하여 수행하였다.

키 워 드

메타데이터 레지스트리, 메타데이터, 표준화, 데이터 요소, 상호 운용성

* 본 연구는 산업자원부 기술표준원의 후원으로 2004년에 수행한 ‘국가산업기술정보 메타데이터 표준화 시범구축’ 사업의 결과를 정리한 것임.

** 연세대학교 전산학과 교수
(Professor, Programming Language Lab, Computer Science, Yonsei University, yknam@dragon.yonsei.ac.kr)

*** 한국과학기술정보연구원 표준화기술지원실 책임연구원
(Principal Researcher, Standardization Technology Support Dept., KISTI, tsseo@kisti.re.kr)

**** 연세대학교 전산학과 대학원
(Programming Language Lab, Computer Science, The Graduate School Yonsei University)

• 논문접수일자 : 2005년 2월 10일

• 게재확정일자 : 2005년 3월 12일

ABSTRACT

The Industrial Metadata Registry(IMR) for standardizing industrial information metadata was implemented and four drafts of metadata sets were developed based on the ISO/IEC 11179. The IMR consists of four individual metadata registries characterized by subject category. Accordingly, the system includes the integrated administration module, and the divisions administration module. The users of the system are divided into committee members and general users. The system has been developed with Oracle 9i and Java JSP over Linux operating system. The system will support the procedure for data sharing and exchanging among users and organizations and improve the availability and manageability of data and provide the standardization of data across the industry.

KEYWORDS

Metadata Registry, Metadata, Standardization, Data Element, Interoperability

1. 서 론

우리나라에서는 정부주도로 많은 예산을 들여 지식정보를 구축하고 있다. 그러나 정보개발에 있어서 표준적이고 정형화된 개발 모델을 따르지 않고, 제작자마다 정보의 형태를 다르게 유통시키고 있음에 따라 동일 분야 데이터베이스들 간의 지식정보의 공유를 어렵게 하여, 결과적으로 추가적인 물적·인적 자원의 낭비를 초래하고 있다.

이러한 문제를 해결하기 위해 학계, 업계에서는 다양한 방법론을 개발하였고, 이를 적용하여 시스템의 통합과 정보공유의 효율을 제고하고자 노력하고 있다(서태설 2002). 국제표준화기구에서 데이터의 교환 및 관리표준화를 담

당하는 ISO/IEC JTC1/SC32(Data Management and Interchange)에서는 데이터의 의미, 구문, 표현을 표준화할 수 있는 프레임워크를 제시하기 위해서 'ISO/IEC 11179 메타데이터 레지스트리'를 제정하였다. 이 표준을 이용하면 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지·관리함으로써 메타데이터의 명세와 의미를 공유할 수 있다. 현재 미국을 비롯한 여러 국가의 주요 프로젝트에서 메타데이터 레지스트리 시스템을 구현하고 있으나, 우리나라에서는 이에 대한 이론적인 연구는 많이 있으나 실제적인 시스템 개발은 거의 없는 실정이다.

따라서 국내의 방대한 산업기술 지식정보들에 대한 표준화를 효과적으로 추진하고 관리할 수 있는 모델을 개발하고, 이를 기반으로 정보

의 정확성 및 신뢰성을 일관성 있게 유지할 수 있는 시스템 개발이 필요하다. 산업기술정보에 있어서도 이러한 문제가 동일하게 대두되고 있어서, 우리나라의 표준을 총괄하고 있는 기술 표준원(Korea Agency for Technology and Standards)에서는 이에 대한 해결방안으로서 산업기술정보에 대한 메타데이터 표준을 만들고 보급하는 사업을 추진하고 있다. 이 사업에서는 ISO/IEC 11179에 기반하여 산업기술정보 메타데이터 레지스트리(IMR: Industrial Metadata Registry)를 구축하고, 주요 정보에 대한 메타데이터 표준안을 개발하였다.

본 논문에서는 산업기술정보 메타데이터 표준화를 ISO/IEC 11179에 따라 수행한 내용을 정리하였다. 4개의 산업기술정보 메타데이터 표준안이 개발되었고, 이러한 메타데이터 표준안의 데이터 요소를 정의하고 표준화 할 수 있는 산업기술정보 메타데이터 레지스트리를 구현한 내용을 소개한다.

2. 관련 연구

메타데이터(metadata)는 일반적으로 데이터에 관한 데이터로서 정보자원의 속성을 기술하는 데이터를 의미한다. 즉, 메타데이터란 실제로 저장하고자 하는 데이터(예를 들면, 비디오, 오디오, 텍스트 등) 자체는 아니지만, 이 데이터와 직접적으로 혹은 간접적으로 연관된 정보를 제공하는 데이터를 나타내는 말이다. 일반적으로 메타데이터는 “데이터에 대한 데이

터”, “정보자원을 설명하는 요소집합” 등으로 정의된다. 이와 같은 메타데이터를 사용하면, 사용자가 원하는 데이터가 맞는가를 확인할 수 있고, 쉽고 빠르게 원하는 데이터를 찾아낼 수 있다. 즉, 데이터를 소유하고 있는 측면에서는 관리의 용이성을, 데이터를 사용하고 있는 측면에서는 검색의 용이성을 보장받을 수 있기 때문에 메타데이터의 필요성이 더욱 높아지고 있다.

인터넷 시대를 맞아 더블린 코어(Dublin Core) 메타데이터 세트가 작성되면서 메타데이터 표준화는 활발히 연구되고 있다. 국제적으로 메타데이터에 대한 표준은 여러 가지 표준이 경쟁적으로 만들어지고 있기 때문에 단순히 정리하기가 쉽지 않다. 현재 개발된 메타데이터로는 그 응용분야에 따라 더블린 코어, MODS, INDECS, ONIX, MPEG21, TV Anytime 등이 있으며, 이외에도 많은 메타데이터 표준이 개발되고 있다(Sally 2004).

메타데이터 국제표준을 다루고 있는 기관은 국제 표준기구인 ISO/IEC JTC1, DCMI(Dublin Core Metadata Initiative), 미국의 회도서관(LC: Library of Congress), <indec> Framework Ltd., TVAF(TV Anytime Forum), OAI(Open Archives Initiative), DIG(Digital Imaging Group) 등이 있으며, 이외에도 많은 기구 또는 단체들이 존재하고 있다(Condron, Tittlemore 2001).

국내에서도 각 기관별로, 그리고 업계별로 나름대로의 메타데이터 표준을 개발하여 사용

하는 추세에 있다. 결과적으로 동일한 정보에 대해서도 다르게 정의되어 구축됨으로써 기관 간의 정보공유에 많은 문제가 야기되고 있다(서태설, 이윤석, 김이관 2002).

이러한 문제를 해결하기 위한 노력 가운데 하나는 모든 형태의 네트워크 자원을 하나의 표준적인 메타데이터로 표현하려는 움직임이 있다. 이는 다양한 정보자원과 메타데이터를 기존 시스템 틀 안에서 통합하여 조직할 수 있는 가장 효율적인 메타데이터를 선정하고, 이를 표준적인 통합 메타데이터 포맷으로 사용하자는 것으로 MARC(Machine Readable Cataloging)를 대표적인 예로 들 수 있다(심경 2003).

각 분야별로 다양한 메타데이터를 사용하는 것으로 인해 발생하는 문제를 해결하고자 하는 두 번째 주장은 다양한 메타데이터 형식과 기술구조를 인정해야 한다는 관점이다. 현재 각기 다른 기술구조를 가진 다양한 메타데이터를 상호 운용성의 입장에서 통합하기 위한 연구가 이루어지고 있으며, RDF(Resource Description Framework)와 같은 시맨틱 웹 기술이 그 예이다. 이 두 가지 방향은 각각의 장단점이 있다. 첫 번째 방법은 수많은 데이터 요소를 표준화 하여야 하기 때문에 사용자가 취사선택하여 사용할 수 있으나, 표준화의 비용이 많이 들고, 표준의 적용에도 전문적인 지식이 필요하여 특정한 이외의 사람들은 사용하기가 어렵다. 두 번째 방법은 다양성을 인정하기 때문에 메타데이터 개발이 많이 이루어지겠지만, 실질적인 적용사례가 적어 검증이 불충분하고

XML 구문이 아닌 영역에는 적합하지 않다.

ISO/IEC 11179 표준에 기반한 메타데이터 레지스트리는 이러한 두 가지 방향을 모두 수용할 수 있는 기반을 제시하고 있어서, 미국 환경청, DCMI 등 여러 기관에서 이러한 방식을 적용하고 있다. 따라서 우리나라도 이러한 추세에 맞추어서 분야별 메타데이터 레지스트리를 구축하여 국가 지식정보의 메타데이터를 표준화해 나가는 것이 바람직할 것이다(한국, 기술표준원 2004).

선진국에서는 이미 메타데이터 레지스트리 표준화 연구와 공유체제 구축에 광범위하고 집중적인 투자를 하고 있다. 메타데이터 레지스트리 표준화 및 구현은 미국 환경청의 EDR(Environmental Data Registry)가 가장 오래되었으며, 미 환경청은 미국의 국가표준화 기구인 ANSI와 함께 이 분야의 국제표준(ISO/IEC 11179)을 주도하고 있고, 교통부의 지능형 교통 시스템(ITS)을 위한 MDR와 국방성의 건강정보지식 베이스(USHIK)를 위한 MDR가 구축되어 있다.

그밖에 유럽에서는 SCHEMAS registry가 구축되어 22개 전문분야로 확산 수용되고 있으며, 호주의 AIHW(The Australian Institute of Health & Welfare)에서도 건강정보를 위한 MDR를 구축하여 운영 중이고, DCMI(Dublin Core Metadata Initiative)는 1999년 12월에 DCMI Registry WG을 구성하여 ISO/IEC 11179에 기반하여 MDR 프로젝트를 진행하였다(황상원 외 2004).

〈표 1〉 ISO/IEC 11179의 구성

Part	규격 명칭	제정년도
1	Framework	1999
2	Classification	2000
3	Registry metamodel and basic attributes	2003
4	Formulation of Data Definitions	1995
5	Naming and identification	1995
6	Registration	1997

국내의 경우 메타데이터 레지스트리 기술에 대한 연구는 2001년 이후에 소프트웨어 컴포넌트 재사용을 위한 MDR 연구(송치양 외 2001), 문헌정보 유통표준화를 위한 MDR 프로토타입 연구가 있었고(김이란, 서태철, 이운석 2002), 최근에는 메타 모델 상호 운용성 연구, R&D 정보유통 표준화를 위한 MDR 연구가 부분적으로 수행되고 있다(유상훈 외 2002; 정동원 외 2003).

3. ISO/IEC 11179의 개요 (ISO 2003)

ISO/IEC 11179는 전체적으로 6개 부분(Parts)으로 구성되어 있다(〈표 1〉 참조). 제1부(Part 1)는 MDR의 프레임워크를 제시하며, 제2부(Part 2)는 데이터 요소의 분류를 다룬다. 제3부(Part 3)는 MDR의 메타모델(meta-model)과 메타데이터 기본 속성을 제시한다. 제4부(Part 4)와 5부(Part 5)에서는 데이터 요소의 명명, 식별, 정의에 대한 지침을 제공한다. 마지막으로 제6부(Part 6)에서는 데이터

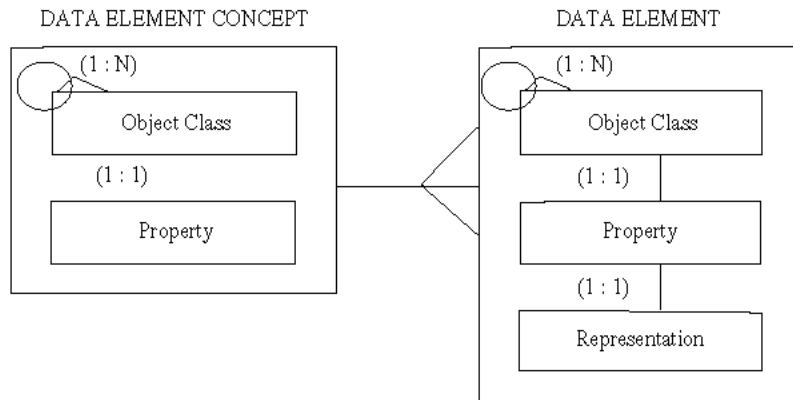
요소 식별자 부여방법과 함께 MDR를 이용하여 데이터 요소를 제안하고 표준상태로 만들어 가는 절차에 대해서 규정하고 있다.

3.1 메타데이터 레지스트리 기본 모델

ISO/IEC 11179의 1부에서는 메타데이터 레지스트리에 의해 관리되는 데이터의 기본단위에 대한 설명을 다룬다. 그 기본 단위는 데이터 요소이며, 데이터 요소는 3가지 구성요소로 이루어진다(〈그림 1〉 참조).

- 객체 클래스(Object Class) : 실세계의 생각, 추상 또는 사물들의 집합으로 명백한 범위와 의미, 그리고 속성, 행위들이 같은 법칙에 의하여 정의된다.
- 특성(Property) : 한 객체 클래스 내의 구성 요소가 가지는 일반적인 특성
- 표현(Representation) : 데이터를 어떻게 표현하는가에 관한 문제로, 값영역(Value Domain), 데이터 타입(Data Type) 등에 해당한다.

ISO/IEC 11179의 3부에서는 공유 데이터의



〈그림 1〉 데이터 요소의 구성요소

관리를 위한 메타 모델을 제시한다. 메타 모델은 의미적인 내용과 분산된 환경하의 사용자들이나 정보처리 시스템 간의 공유되는 데이터 요소의 구문을 위한 표준과 안내를 제공하고 있다. 이 표준은 메타데이터 레지스트리의 구조를 개념적 메타 모델의 형태로 명시하고 있다. 메타데이터 레지스트리의 구조는 관리 및 식별(Administration and Identification), 명명 및 정의(Naming and Definition), 분류(Classification), 관리항목(Administered Item) 등 4개의 영역으로 구성된다.

관리 및 식별 영역에서는 데이터 요소관리를 위한 메타데이터 항목으로 관리항목의 등록자 정보, 식별정보, 제안자 정보, 담당자 정보, 참조정보, 등록/생성 일자, 유효일자, 관련 comment, note 정보 등을 다룬다.

명명 정의 영역에서는 관리항목이 활용되는 주제 분야, 사용언어의 명시, 동의어 등 다양한 표현방법 지원, 언어별 다양한 의미의 사

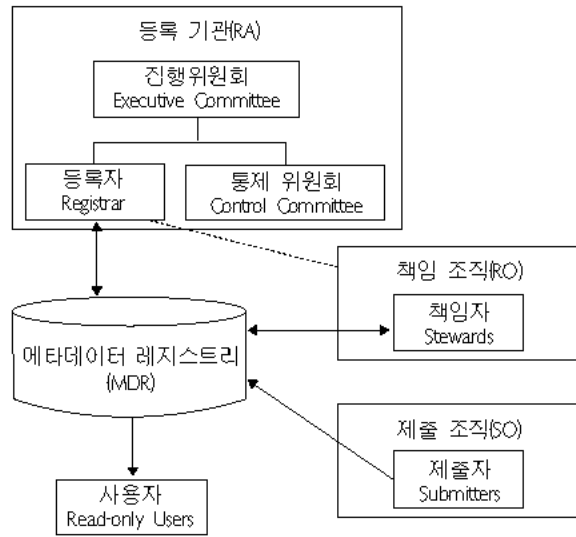
용 지원 등을 다룬다.

분류 부분은 분류 스키마와 분류 스키마 내에 존재하는 저장소 구성요소를 관리하는 데에 사용되는데, 분류체계의 유형 정의, 분류체계를 구성하는 항목, 항목의 관계정의를 다룬다.

관리항목으로 대표적인 것은 데이터 요소, 데이터 요소 개념, 개념 영역, 값영역이 대표적이다. 데이터 요소 개념관리는 데이터 요소의 의미에 해당하는 개념영역에 대한 관리이고, 데이터 요소에 대한 관리는 데이터 요소 자체에 대한 관리이다. 값영역과 개념영역의 관리는 실질적인 표현에 해당하는 부분과 관련된다.

3. 2 메타데이터의 등록

ISO/IEC 11179 제6부인 '데이터 요소의 등록'에서는 등록자(registrar)를 통하여 데이터



〈그림 2〉 MDR를 이용하는 등록활동 주체

요소를 등록하고, 검증과 인증을 통하여 데이터 표준화를 지원한다. 데이터 요소를 표준화하는 일은 메타데이터 레지스트리를 중심으로 여러 이해 당사자가 관련이 되어서 이루어진다. 따라서 각 이해 당사자들의 역할과 책임을 분명히 하는 것이 필요하다. 메타데이터 등록 프로세스와 관련된 등록활동 주체(RAB: registration acting body)의 구성은 〈그림 2〉와 같다.

MDR는 메타데이터의 등록, 수정, 삭제, 검색, 적재 및 검증, 인증, 표준화, 상태 검색 과정의 투표 기능을 가진다. 각 전문분야별 기관이 제안하는 메타데이터를 표준화된 데이터 요소로 등록하기 위해서는 위원회의 의결 및 승인과정을 거쳐야 한다. 데이터 요소의 제안에서 표준화 과정까지의 담당자 및 역할은 등급

에 따라 다음과 같이 분류한다. 메타데이터의 데이터 요소를 등록, 수정 및 삭제하는 제출자(Submitter), 제출자에 의해 제출된 메타데이터를 검증하는 권한을 갖는 책임자(Steward), 책임자의 검증이 완료된 메타데이터를 인증하는 등록자(Registrar), 등록자가 인증한 메타데이터를 표준화하는 통제위원회(Control Committee)가 있다. 이용자(User)는 MDR로부터 표준화된 데이터 요소를 검색 및 다운로드할 수 있다. 모든 위원은 가입시 개인정보와 역할에 맞는 희망등급을 등록하며, 등록된 위원은 분야별 관리자의 인증 후 MDR를 이용할 수 있다. 그렇지만, 이용자는 회원가입 절차 없이 MDR를 이용할 수 있다.

4. 산업기술정보 메타데이터 표준안 개발

산업기술정보란 기업이나 연구소가 산업기술을 연구하고 제품을 개발하는 과정에서 발생하는 정보를 말하는데, 이러한 정보에는 산업인력정보, 연구정보, 특허정보, 기업정보, 상품정보, 규격정보, 기자재정보, 산업무역 통계정보, 조달정보, 법령정보, 기술이전정보 등이 포함된다. 본 연구에서는 산업기술정보 중에서 규격, 연구정보, 특허, 기업·상품 정보에 대해서 시범적으로 메타데이터 표준안을 개발하였다.

메타데이터 표준안의 개발절차는 우선 해당 분야의 정보설계 경험이 있는 사람으로 하여금, 해당 정보의 데이터 요소를 추출하도록 한 후, ISO/IEC 11179의 3 5부의 지침에 따라 요소를 정의하도록 하였다. 그리고 관련 전문가들의 검토를 거쳐서 표준안을 완성하였다.

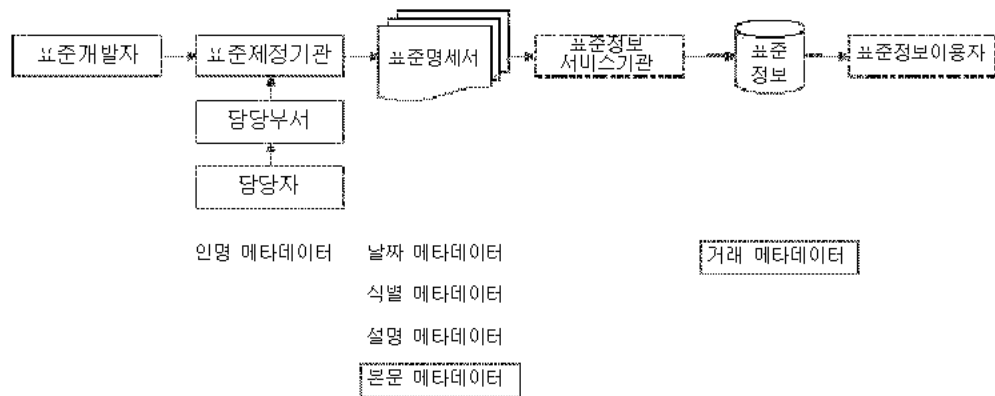
4.1 데이터 요소의 선정

메타데이터의 선정은 기존에 나와 있는 관련 표준과 데이터베이스의 항목명을 참고하여 이루어졌다. <표 2>에 각 대상 정보별 데이터 요소추출 범위와 참고자원을 정리하였다. 또한 각 정보에 대한 데이터 모델을 작성하여 각 요소의 정당성을 확인하였다. <그림 3>은 규격정보의 데이터 모델과 관련 메타데이터 종류를 나타낸다.

대상으로 하는 4개의 산업기술정보는 대부분 인력 및 기관정보, 문서의 본문 내용을 기술하는 정보, 날짜와 관련된 정보, 식별성 정보, 요약 및 설명정보, 거래와 관련된 정보 등으로 구성된다. 본 연구에서는 공통적이고 기본적인 정보를 중심으로 메타데이터 요소를 추출한 결과 인명, 날짜, 식별, 설명, 본문, 거래 등의 6개 유형의 메타데이터를 중심으로 정의하였다. <표 3>에는 각 대상 정보별 메타데이터 유형과

<표 2> 대상범위 및 참고자원

대상 정보	요소추출 범위	참고자원
규격정보	KS 및 ISO 표준문서를 위한 서지 및 원문 요소	ISO/IEC Directive Part 2 KS A 0001:2003
연구정보	정부 및 공공기관에서 발주하는 연구과 제 및 보고서의 서지요소	CERIF Model, INSPEC XML DTD 국내 연구보고서 DBs
특허정보	한국 특허를 포함한 전세계 특허의 서지 및 권리 요소	INID Code & 특허 명세서 Dublin Core Metadata Set RFC 1807
기업·상품정보	기업 및 상품의 식별, 설명, 거래 요소	GDAS KS X 6722-1 G2B 정부 표준



〈그림 3〉 규격정보의 데이터 모델과 메타데이터

〈표 3〉 메타데이터 유형별 데이터 요소 수

대상 정보	요소 수						합 계
	인명	날짜	식별	설명	본문	거래	
규격정보	2	6	3	18	16	4	49
연구정보	12	2	10	24			48
특허정보	8	14	7	19			48
기업·상품정보			6	19			25
합 계	22	22	26	80	16	4	170

추출된 데이터 요소 항목수를 나타내었다. 4개 대상정보에 대해서 모두 170개의 데이터 요소를 선정하여 정의하였다.

4. 2 데이터 요소의 정의

데이터 요소를 정의하기 위한 기본 속성은 ISO/IEC 11179 Part 3에서 상세히 제시하고 있다. 그 중에서 필수 속성과 산업기술정보에 적합한 속성을 우선적으로 선정하였으며, 표준에서는 제시되어 있지 않지만, 필요성이 있는 속성을 추가하여 모두 31개의 메타데이터 속성

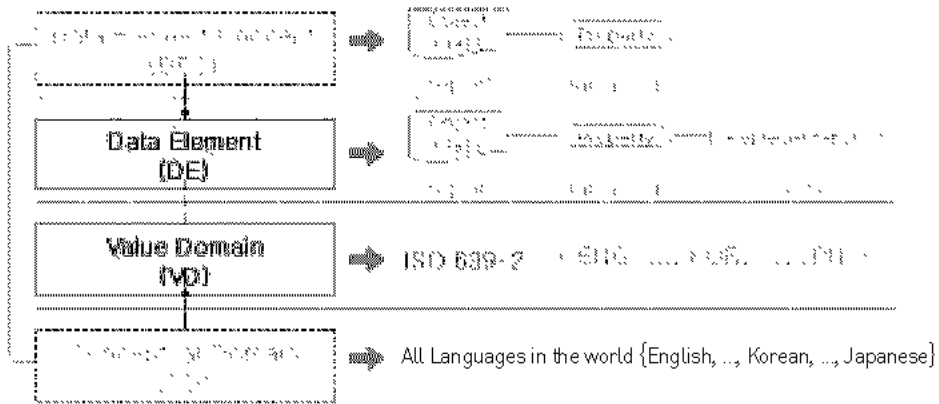
을 정의하도록 하였다. 각 데이터 요소를 정의할 때 데이터 요소 이름, 식별자, 정의에 해당되는 부분은 ISO/IEC 11179의 Part 4와 5의 지침을 충실하게 따르도록 하였다.

〈그림 4〉는 ‘연구보고서 언어 코드(report language code)’에 대해서 정의한 사례를 나타낸다. ISO/IEC 11179의 Part 5에 따르면 데이터 요소(data element)는 객체 클래스(object class)와 특성(property)이 결합된 데이터 요소개념(data element concept)에 표현(representation) 부분이 결합되어서 이루어진다(〈그림 1〉 참조). 예를 들어, ‘연구보고

서_언어_코드(report_language_code)의 경우, 이것은 연구보고서(report)라는 객체와 언어(language)라는 특성이 결합된 데이터 요소 개념에 코드(code)라는 표현 부분이 결합되어서 이루어진다. 데이터 요소 중에서 표현 부분이 특성과 중복되는 경우는 표현 부분을 생략할 수 있다. 이 경우는 데이터 요소개념과 데이터 요소가 같게 된다. 예를 들어, '기업_등록번호'의 경우 원래는 '기업_등록번호_번호'가 맞

으나 번호라는 말이 중복되므로 생략해서 표현한다.

또한 데이터 요소에 상응하는 값영역(Value Domain)은 이 세상의 모든 언어라는 개념 영역(conceptual domain)을 코드화 한 ISO 639-2가 허용값(permissible value)으로 지정될 수 있다. 다른 모든 데이터 요소도 이와 같이 동일한 원칙으로 명명되고 정의되었다.



데이터 요소 식별자	RDDE040100
데이터 요소 이름	report language code
데이터 요소 한글이름	연구보고서 언어 코드
정의	연구보고서를 작성하는데 사용된 언어
정의 언어	한국어
정의 참고정보원	문헌정보학 용어사전
분류체계명	IMR 주제 분류
분류체계 유형명	메타데이터 용어집
분류체계 항목유형명	용어
분류체계 항목값	설명
개념영역	언어
값의미 기술	작성언어
허용값	ISO 639-2
값영역	언어코드
값영역 데이터 유형	문자열 영문

데이터유형	문자열 영문
데이터유형 참조스킴	
최소문자수	3
최대문자수	3
데이터요소개념 이름	연구보고서 언어
객체 클래스 이름	report
객체 클래스 한정자	문자열
속성 이름	language
속성 한정자	문자열
표현클래스	code
표현클래스 한정자	문자열
주제기술	연구보고서
주제명	언어
버전	1.0
주석	
사례	KOR

〈그림 4〉 데이터 요소정의 사례

4. 3 공통 메타데이터

공통 메타데이터란 산업기술정보를 표현하기 위한 메타데이터, 특히 규격정보, 연구정보, 특허정보, 기업·상품 정보에서 공통적으로 활용할 수 있는 개념 영역에 해당하는 관리항목에 대해서 고려하여 개발되었다. 따라서 다른 분야에 적용할 때에는 그 분야에 맞게 수정되어야만 한다.

예를 들면, 언어라는 개념 영역의 경우, 연구정보에서는 '연구보고서_언어_코드', 특허정보의 경우, '출원_최초언어_코드'로 다르게 나타나지만, 값영역에서는 동일한 언어 코드(ISO 639 2)를 사용하게 된다. 이와 같이 주제분야에 따라서는 다른 데이터 요소 이름을 가지지만, 값영역은 같은 경우에 대해서 공통 메타데이터를 규정함으로써 값 표현을 통일할 수 있다는 이유에서 공통 메타데이터를 정의하였다. 이러한 것으로는 가부, 언어, 날짜, 이름, 우편

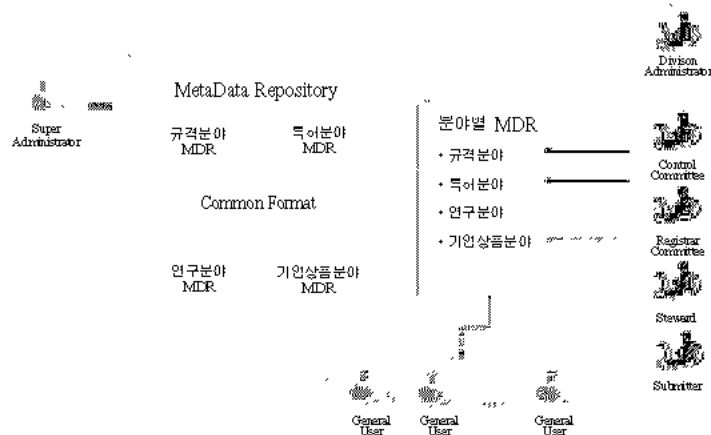
번호, 통화 등을 들 수 있다.

이 부분에 대해서는 데이터 요소간의 의미 공유라는 더 넓은 차원에서의 더 깊이 있는 연구가 필요할 것으로 사료된다.

5. 산업기술정보 메타데이터 레지스트리(IMR) 개발결과

5. 1 시스템 개요

메타데이터 관리 표준 모델 및 데이터 요소 관리 절차를 설계, 개발하였다. 개발된 시스템은 산업기술정보 분야의 데이터 공유 및 교환을 위한 표준절차를 제시하며, 정보의 활용성, 관리 편의성 및 표준 데이터 모델을 제공하는 시스템이다. 이 레지스트리는 각 분야를 총괄해서 관리하는 통합관리자용 관리 부분, 분야별 시스템을 관리하는 분야별 관리자 모듈, 분야별 위원회를 위한 모듈 및 분야별 일반 사용자



<그림 5> IMR 아키텍처

〈표 4〉 사용자별 기능과 권한

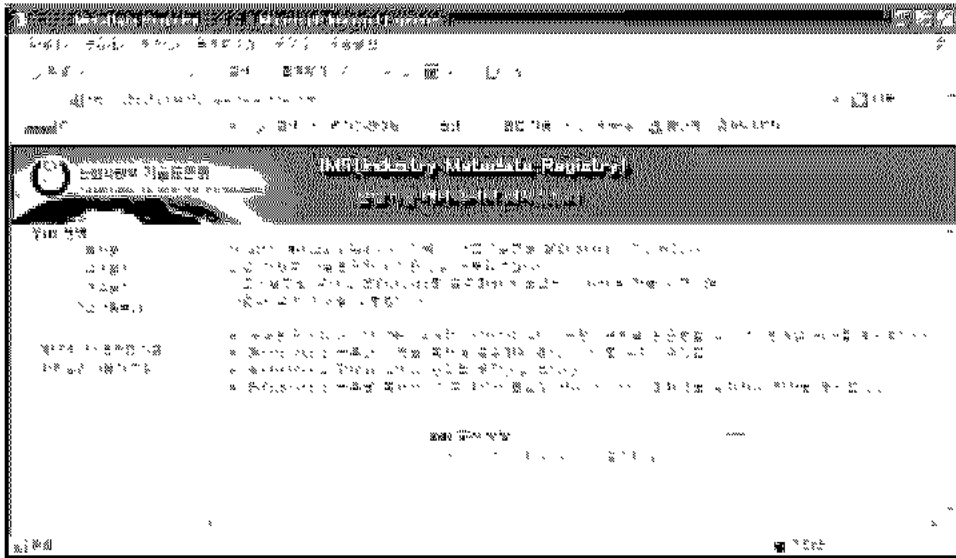
구 분	통합관리자	분야관리자	제안자	책임자	등록위원	통제위원	이용자
회원정보 수정	○	○	○	○	○	○	○
회원정보 현황조회		○	○	○	○	○	
회원가입 신청 처리		○					
메타데이터 검색	○	○	○	○	○	○	○(비로그인)
스크립트 생성							○
메타데이터 입력			○				
메타데이터 수정			○	○	○	○	
메타데이터 삭제			○				
메타데이터 상태검색		○	○	○	○	○	
투표기준 관리		○					
투표기준 조회			○	○	○	○	
투표					○	○	
통계 보기	○						
메타데이터 분야관리	○						

자를 위한 모듈로 구성되었다. 데이터 요소의 표준화를 위한 절차는 데이터 요소를 구분하는 영역을 별도의 화면으로 구성하여 등록 및 검색의 효율을 기하고, 투표과정에서 데이터 요소의 상태 흐름에 따라 자동적으로 표준화가 유도되도록 데이터 요소의 등록절차를 구현하였다. 관리자를 위한 다양한 통계기능과 위원들 간의 상태정보가 제공되며, 일반 사용자를 위한 다양한 데이터베이스 스크립트를 제공하여 이용자 편의성을 높였다.

IMR는 〈그림 5〉에서 보는 것처럼 4개의 분야별 MDR로 나누어서 개발하였다. 따라서 관리자는 통합관리자와 분야별 관리자로 나누어지며, 사용자에는 제안자(Submitter), 책임자(Steward), 등록위원(Registrar), 통제위원회(Control Committee), 이용자(User)가 존재

한다. 〈표 4〉에는 각 사용자별 기능과 권한을 나타내었다.

산업기술정보 MDR의 초기화면은 〈그림 6〉과 같다. 각 분야별 관리자는 초기화면 좌측의 'MDR 분야선택' 메뉴에서 접근하려는 MDR 분야를 선택하여 해당 분야의 MDR를 관리하나, 통합관리자는 전체 시스템을 관리하며, 새로운 분야의 MDR를 시스템에 추가하거나, 현재 관리하고 있는 MDR의 분야명을 변경 또는, 더 이상 사용하지 않는 MDR 분야를 삭제하는 권한을 갖는다. 또한 통합관리자는 산업기술정보 MDR에 등록된 모든 회원에 대해 상단의 분야 및 등급 선택 리스트에서 분야별, 등급별 혹은 전체 MDR 회원 현황을 구분하여 조회할 수 있다. 이러한 정보는 각 분야별 MDR에서 등록, 관리되므로 별도의 회원관리 기능은 수행



〈그림 6〉 산업기술정보 메타데이터 레지스트리 초기화면

하지 않는다.

각 분야별 관리자는 자신이 관리하는 분야의 회원에 대한 권한만을 갖는다. 분야별 모든 회원은 개인정보와 희망등급 등을 입력하여 회원으로 등록하며, 등록된 회원은 분야별 관리자의 승인을 얻어 시스템에 참여한다.

그렇지만, 일반 사용자는 회원가입 절차를 거치지 않더라도 MDR 시스템을 이용할 수 있다. 그러나 자신들의 데이터베이스 생성을 위해 MDR의 데이터 요소를 이용할 경우에는 분야별로 사용자 등록을 거쳐야 한다. 분야별 관리자는 회원에 대한 변경 및 탈퇴권한도 갖는다.

제안자는 자신이 입력한 데이터 요소를 검색, 수정·삭제할 수 있다. 다만, 데이터 요소의 식별자(Identifier)는 수정할 수 없으며, 식별자의 변경은 등록 데이터 요소 전체의 의미

변경을 뜻하므로, 삭제 후 새로운 데이터 요소로 등록해야 한다.

책임자, 등록위원, 통제위원은 메타데이터 검색, 수정, 상태검색만 할 수 있다. 등록위원과 통제위원은 투표 기준 조회와 투표를 할 수 있다.

5. 2 데이터 요소의 입력

MDR에서 공유할 새로운 데이터 요소는 각 분야별 '제안자(Submitter)'가 등록하며, 데이터 요소는 6단계로 나누어 단계적으로 입력하도록 하였다. 즉, ①관리(Administration), ②분류(Classification), ③개념 및 값영역(Conceptual and Value Domain), ④데이터 요소개념(Data Element Concept), ⑤명명 및

데이터 요소 이름(name)	<input type="text"/>	등록검색
데이터 요소 한글이름(korean name)	<input type="text"/>	
데이터 식별자(data identifier)	STDE0050100	
버전(version)	1.0.0	
등록상태(registration status)	recorded	
관리상태	<input type="text"/>	
생성일자	20041227	
최종수정일자	20041227	
유효일자	20101212	
마감일자	20101212	
변경내용	<input type="text"/>	
관리 노트	<input type="text"/>	
추가설명	<input type="text"/>	
이하결 사항	<input type="text"/>	
출처	<input type="text"/>	
책임기관명(responsible organization name)	산업자원부 기술표준원	
제출기관명(submitting organization name)	<input type="text"/>	
국제지정 코드	<input type="text"/>	
조직부서 식별자	<input type="text"/>	
조직부서 식별자 출처	<input type="text"/>	
등록기관 식별자	<input type="text"/>	

〈그림 7〉 관리부분의 입력화면

식별(Naming and Identification), ⑥데이터 요소(Data Element)의 6개 부분으로 구성하였다. 〈그림 7〉은 관리부분의 입력화면을 나타낸다.

5. 3 데이터 요소의 상태관리

데이터 요소의 상태관리는 데이터 요소의 등록 여부보다 더 복잡하므로 등록하는 데이터의 품질을 높이기 위해 데이터 흐름의 상태가 관리되어야 한다. 각 데이터 요소는 위원들의 투표 과정에 따라 ‘등록(Recorded)’, ‘검증(Qualified)’, ‘인증(Certified)’, ‘표준(Standard)’으로 나누어진다. 각 단계 사이에는 관리를 위한 상태수준이 존재하며, 이러한 상태수준을 관리단계라고도 한다. 메타데이터

레지스트리에 저장되는 메타데이터 생명주기는 다음과 같다.

- 등록(Recorded) : 데이터 요소의 내용이 ISO 11179의 각 부에서 제시하는 품질 요구 사항에 부합되지 않는 단계로 제안자에 의해 메타데이터로 입력된 상태
- 검증(Qualified) : 데이터 요소의 내용이 ISO 11179의 각 부에서 제시하는 품질 요구 사항에 부합되는지를 검증하는 단계로 관리자에 의해 검증된 상태
- 인증(Certified) : 데이터 요소의 내용이 ISO 11179의 각 부에서 제시하는 품질 요구 사항에 부합된 단계로 등록위원회에 의해 인증된 상태
- 표준(Standard) : 등록기관에 의하여 인증된 데이터 요소가 만들어진 단계로 제어위원

의 투표가 완료되어 표준화로 채택된 상태

5. 4 데이터 요소 투표

산업기술정보 메타데이터 레지스트리의 사용자는 원하는 데이터 요소가 없을 경우 일정한 양식에 의해 제안자에게 이메일을 통해 원하는 데이터 요소의 등록을 요청하며, 제안자는 데이터 요소를 등록한다. 등록된 데이터 요소는 등록상태(status)가 'Recorded'로 되며, 이 데이터 요소는 다른 위원들의 투표를 거쳐 최종적으로 표준화 데이터 요소가 된다. 데이터 요소 등록 절차의 유연성을 위해 분야별 관리자는 각 위원별 투표 기준표를 화면에서 생성한다. '투표 인증률'은 해당 위원의 몇 퍼센트가 찬성하면 다음 단계로 상태를 변경할 것인지에 대한 기준율이며, '투표마감기간'은 각 위원의 이전 등급의 위원으로부터 투표가 완료되어 등록상태가 변경된 일자로부터 며칠간을 해당 위원의 투표 기간으로 정할 것인지 기간을 입력한다. 예를 들면, 등록 위원들의 투표 인증률 이상을 얻은 데이터 요소가 'Qualified' 상태에서 'Certified' 상태로 바뀐 일자에 다음 등급인 제어위원의 투표 기간을 더하여 제어위원의 투표 마감일을 결정한다. 이러한 과정을 거쳐 최종적으로 표준화된 데이터 요소는 분야별 관리자의 적재 작업을 통해 MDR에 저장하여 사용자에게 제공한다. 만약 투표 인증률을 통과하지 못했거나, 투표 마감일까지도 투표가 안된 데이터

요소는 시스템에서 삭제되며, 삭제된 데이터 요소는 등록된 사용자가 다시 등록을 원한다면, 제안자에게 관련된 내용을 이메일로 보내 위와 같은 처리절차를 통해 다시 투표가 이루어지게 한다.

모든 위원의 '투표상태 조회' 화면은 동일하나, 각 위원별로 제공되는 데이터 요소의 'status'가 등급에 따라 다르게 표시된다. 즉, 데이터 요소별로 관리자에게는 제안자가 등록한 'Recorded' 상태가, 등록위원에게는 관리자가 인증한 'Qualified' 상태가, 제어위원에게는 등록위원이 검증한 'Certified' 상태의 데이터 요소가 조회된다.

데이터 요소의 투표는 'VOTE' 버튼을 클릭하면, 데이터 요소별로 투표를 실시할 수 있는 항목이 나타나며, 이 또한 모든 위원에게 동일한 화면이 제공되지만, 각 위원별로 관리 상태가 다른 데이터 요소를 투표하도록 구성하였다. 관리자가 실시할 수 있는 투표상태는 'Qualified', 'Retired', 'Superseded'의 3가지이며, 'Qualified'로 투표한 데이터 요소는 다음 등급인 등록위원의 투표대상으로 상태가 변경된다. 모든 위원들에게 공통으로 적용되는 반대투표 상태가 'Retired', 'Superseded'이며, 'Retired'로 투표할 경우 더 이상 시스템에서 제공하지 않는 데이터 요소가 된다. 'Superseded'로 투표할 경우 'Comment' 항목에 자신의 의견을 입력하여 모든 위원들에게 데이터 요소의 재검토를 제안할 수 있으며, 위원회의 의결을 거쳐 데이터

요소의 재사용 여부를 결정한다. 투표가 완료된 데이터 요소는 각 위원들의 투표목록에서 제외된다. 등록위원이 실시할 수 있는 투표하는 상태는 'Certified', 'Retired', 'Superseded'의 3가지이며, 'Certified'로 투표한 데이터 요소는 다음 등급인 제어위원의 투표대상으로 상태가 변경된다. 제어위원이 실시할 수 있는 투표상태는 'Standard', 'Retired', 'Superseded'의 3가지이며, 'Standard'로 투표한 데이터 요소는 분야별 관리자에게 전달된다.

5. 5 통계 관리

통합관리자, 분야별 관리자, 위원회 및 일반 사용자에게 제공되는 통계정보는 각 등급에 맞도록 구성하였다. 통합관리자에게는 모든 분야에 대한 '분야별/등급별 위원회 현황', '일자별/위원별 투표건수', '위원별/데이터 요소별 투표현황' 등의 위원회와 투표상태를 관리할 수 있는 통계를 제공한다. 분야별 관리자에게는 통합관리자의 통계항목 중 해당 분야에 국한된 정보를 제공한다. 분야별 위원회에게는 '투표 기준표 조회가 제공되며, 일반 사용자에게는 '데이터 요소 및 DB 스크립트 생성건수'에 관한 통계가 제공된다.

5. 6 스키마 설계 및 생성

MDR 시스템이 제공하는 모든 데이터 요소

는 이 시스템을 사용하는 타 기관에 대해 표준화된 메타데이터이다. 각 기관은 MDR 시스템에서 제공되는 데이터 요소를 검색한 후 원하는 요소만을 선택하여 자신들의 스키마 세트로 설계할 수 있다. 사용자 자신들의 데이터베이스를 구축하기 위해 MDR 시스템으로부터 데이터 요소를 이용할 경우는 사용자의 로그인 정보를 입력해야 한다.

사용자가 로그인 과정을 거치면 MDR가 보유하고 있는 데이터 요소들이 제공되며, 'DB 선택' 창을 통해 자신들의 환경에 맞는 상용 데이터베이스를 선택한 후, MDR가 제공하는 Oracle, mySQL, MSSql, Sybase 등의 스크립트를 이용해 각 기관의 시스템에 적합한 데이터베이스를 생성한다. 만약 제공되는 데이터 요소가 없다면, 각 분야별 제안자에게 새로운 데이터 요소를 제안하고, 앞서 기술된 데이터 요소의 등록 및 투표과정을 통해 MDR 시스템에 반영시킬 수 있다.

6. 개발결과 고찰

본 연구에서는 국제표준인 ISO/IEC 11179에 충실하도록 메타데이터 표준안 개발 및 메타데이터 레지스트리 구축을 수행하였다. 이 표준에 따라 메타데이터 레지스트리를 구현한 외국의 사례가 있기는 하지만, 시스템의 소스를 공개하고 있는 곳이 없어서 이번 개발은 최초개발이나 다름없는 개발이라고 할 수 있다. 특히 외국의 사이트들은 대부분 이미 개발된

메타데이터의 검색과 활용에만 초점이 맞추어져 있지만, 이번에 개발된 산업기술정보 메타데이터 레지스트리는 ISO/IEC 11179의 제6부에서 제시하고 있는 메타데이터 표준화 절차를 충실하게 구현하였다는 면에서 특징이 있다. 따라서 아직까지 불충분한 부분이 존재한다. 그것은 표준안 자체가 실제 시스템을 고려해서 만들어지지 않은 부분과 또한 적용주제 분야에 따른 차이와 같은 문제에서 기인하는 것이다.

제일 먼저 이 표준은 영어를 사용하는 국가를 중심으로 만들어졌기 때문에 문화적, 언어적 차이에서 오는 문제가 발생하였다. 예를 들면, 영어에서는 한 단어로 표현되는 것이 한글에서는 두 개의 단어로 분리되는 경우나, 명명 원칙을 따랐을 경우에 표현이 이상한 경우 등을 들 수 있다.

다음으로는 여기서는 분야별로 MDR를 나누었는데, 이것을 통합적으로 활용할 수 있는 가장 좋은 방법을 찾지 못하였다는 것이다. 이것은 현재의 표준에서는 해결할 수 없는 상태이다. 향후 이에 대한 연구가 더 필요할 것으로 사료된다.

또한 MDR를 충분히 이해하지 못하는 이용자의 경우, 이 시스템을 사용하기가 매우 불편할 것이다. 따라서 이용자 수준에 따라서 사용이 편리한 인터페이스의 개발도 중요한 과제가 될 것이다.

이번 개발에서 특기할 만한 것은 Oracle, mySQL, MSSql, Sybase 등의 대표적인

DBMS의 스크립트를 자동으로 생성할 수 있게 한 것이다. 이것은 ISO/IEC 11179에서 제공하고 있는 부분이 아니지만, 이용자 편의성을 크게 제고한 개발이라 할 수 있다.

이번 연구에서는 산업기술정보 중에서 규격, 연구정보, 특허, 기업·상품 정보에 대해서만 메타데이터 표준안을 개발하여 등록하였지만, 향후 전 산업분야로 확대하여 등록하게 되면 정보개발시 DB 설계단계를 크게 효율화할 수 있으며, 이를 널리 보급함으로써 산업기술정보의 공유를 크게 확산시킬 수 있을 것이다.

7. 결 론

본 연구에서는 ISO/IEC 11179에 따라서 산업기술정보에 대한 메타데이터 표준안 개발 및 메타데이터 레지스트리 개발을 수행하였다.

메타데이터 표준안 개발의 경우 국내에서는 처음으로 ISO/IEC 11179 표준에 따라서 데이터 요소를 정의하였다는 의미를 갖는다. 이 과정에서 ISO/IEC 11179 표준을 국내에 적용하는데 검토하여야 하는 지식을 축적할 수 있었다. 또한 여기서 개발된 4개의 메타데이터 표준안은 타분야에서 메타데이터 표준을 개발할 때 좋은 참고자료가 될 수 있을 것이다.

산업기술정보 메타데이터 레지스트리는 4개의 분야별 MDR로 구성되도록 구축하였다. 그에 따라 통합관리자 모듈이 개발되었다. 통합관리자 모듈은 각 분야의 메타데이터 레지스트리를 정의하고, 전 분야의 회원 및 통제를 총

팔해서 관리하는 기능이며, 분야별 관리자 모듈은 분야별로 가입된 회원관리, 메타데이터의 시스템 적재 및 위원회의 투표기준을 정의하는 등 분야별로 메타데이터 레지스트리 시스템을 관리하는 기능이다. 각 분야별 위원회 모듈은 데이터 요소의 등록과 투표과정을 수행하여 새로운 데이터 요소를 메타데이터 레지스트리에 제공하는 기능이며, 일반사용자 모듈은 산업기술정보 메타데이터 레지스트리를 분야별로 검색하여, 자신들의 데이터베이스를 구성하거나, 새로운 데이터 요소를 분야별 위원회에 이메일로 제안하는 기능을 갖는다. 데이터 요소의 표준화를 위한 절차는 데이터 요소를 구분하는 영역을 별도의 화면으로 구성하여 등록 및 검색의 효율성을 높이고, 데이터 요소의 상태흐름에 따라 자동적으로 표준화를 유도할 수 있도록 데이터 요소의 투표과정에서 구현하였다. 관리자를 위한 다양한 통계기능과 위원들을 위한 데이터 요소의 상태정보가 제공되며, 일반사용자를 위한 다양한 데이터베이스 스크립트를 제공하여 시스템 사용률을 높이도록 유도하였다.

본 시스템은 ORACLE 데이터베이스를 기반으로 Java JSP를 이용하여 리눅스 상에서 개발하였다. 향후 과제로서는 본 연구에서 개발된 시스템과 다른 정보 분야의 메타데이터 레지스트리를 연계하는 작업이 필요할 것이며, 본 논문에서는 산업기술정보 메타데이터 레지스트리를 구축하여 타 시스템과의 유연한 인터페이스가 가능한 기반을 마련하였다.

향후 연구로서는 메타데이터 온톨로지 연구, 통합검색기능의 보강과, 사용자 인터페이스의 개선, XML 스키마를 지원하는 메타데이터 레지스트리의 개발 등이 수행되어야 할 것으로 사료된다.

참고문헌

- 김이란, 서태설, 이운석. 2002. ISO/IEC 11179응용: 문헌정보 메타데이터 관리 표준화. [서울: KISTT].
- 서태설, 이운석, 김이란. 2002. 21세기 인터넷 시대의 표준과 기술. [서울: 한국과학기술정보연구원].
- 서태설. 2002. 지식정보 인프라 표준화의 중요성과 전망. 『지식정보인프라』, 9: 2 7.
- 송치양, 임성빈, 백두권, 김철홍. 2001. 컴포넌트 유통환경을 위한 컴포넌트 메타데이터 레지스트리 구축: C_MDR. 『정보과학회 논문지 : 컴퓨팅의 실제』, 7(6): 614 629.
- 심경. 2003. 메타데이터 통합 방안. 『한국도서관·정보학회지』, 34(3): 169 192.
- 유상훈, 정동원, 신동길, 서태설, 백두권. 2002. 데이터 우선순위 기반의 점진적인 데이터 통합을 위한 메타 모델 설계. 『한국정보과학회 학술발표회 논문집』: 175 177.
- 정동원, 신동길, 정은주, 이정욱, 서태설, 백두권. 2003. 데이터 가시성 기반의 점진

- 적 메타데이터 레지스트리 구축 방법론. 『정보과학회논문지』, 9(6): 610-622.
- 한국. 기술표준원. 2004. 국가산업기술정보 메타데이터 표준화 시범구축 사업: 분야별 메타데이터 표준안.
- 황상원, 최귀자, 김재곤, 송인석, 서태설, 박동인, 남영광. 2004. 산업기술정보 표준을 위한 ISO 11179 메타데이터 레지스트리 개발. 『KOSTI Workshop 2004』. (대전): 한국과학기술정보연구원.
- Condron, Lyn and Cecilia Piccolo Tittmore. 2001. "Metadata Standards for Library Catalogers". 33(1): 17-20.
- Sally H. McCallum. 2004. "An Introduction to the Metadata Object Description Schema (MODS)". *Library Hi Tech*, 22(1): 82-88.
- ISO. 2003. Information Technology metadata Registry. ISO/IEC 11179.