

# 온톨로지 기반 메타데이터 명명 규칙에 관한 연구\*

## A Study on the Naming Rules of Metadata based on Ontology

고 영 만(Young Man Ko)\*\*

서 태 설(Tae-Sul Seo)\*\*\*

### 초 록

본 연구에서는 정보자원의 메타데이터 작성시 메타데이터 간 의미의 일관성을 유지하기 위한 메타데이터 명명 방법론과 이를 실제 분야에 적용할 수 있는 메타데이터 명명 규칙의 실험적 모형을 제시하였다. 이를 위해 우선적으로 ISO/IEC 11179에서 제시하고 있는 메타데이터 레지스트리 메타모형과 데이터의 기본 속성 및 개념을 논의하였으며, 이러한 논의를 토대로 객체용어(object term)와 속성용어(property term) 및 표현(presentation)에 관한 명명 규칙의 실제 적용 사례를 제시하였다. 객체용어의 생성은 자료유형의 엔터티-관계(E-R) 모형에 근거한 휴리스틱 분석을 통해 이루어졌으며, 속성용어의 명명은 더블링크어의 메타데이터 셋을 기반으로, 표현은 SHOE 1.0 버전을 기반으로 하였다.

### ABSTRACT

To build the consistency among different metadata systems and to increase the interoperability of that systems even among different domains, naming rules and glossaries for the data elements are necessary. This study provides discussion of naming and identification of the data element concept, data element, conceptual domain, value domain, and its meta model. This study also describes example naming conventions based on ontology derived from the combination with object, properties, and representation of data elements. The naming principles and rules described in this study use E-R analysis, DC metadata set, and SHOE 1.0 as an example of the scientific documents. This study would be a guideline to build the naming rules of metadata based on ontology in various domains.

키워드: 메타데이터, 메타데이터 레지스트리, 데이터 요소, 메타 모델, 온톨로지, 명명 규칙, 상호운영성  
metadata, metadata registry, data element, meta model, ontology, naming rule, interoperability

---

\* 본 연구는 2005년도 정보통신기술협회의 지원에 의한 "표준개발 연구"의 일부로 수행되었음.

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

\*\*\* 한국과학기술정보연구원 표준화기술지원실 책임연구원(tsseo@kisti.re.kr)

■ 논문접수일자 : 2005년 11월 11일

■ 게재확정일자 : 2005년 12월 9일

## 1. 서론

전통적으로 온톨로지란 존재와 존재자(存在者 : 존재하는 것)의 본성을 연구하는 형이상학의 한 부분으로서 세상의 구성 요소에 대한 명확한 이해를 얻고자 하는 철학의 연구 분야이다. 오늘날 정보자원 관리와 관련해서 차용하고 있는 온톨로지(ontology)라는 용어는 '사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상에 대하여 공유하는 개념을 명확하고 명시적으로 정의하고 규정하는 것'으로 그 의미가 전이되어 사용되고 있으며, 특히 컴퓨터 분야에서는 '공유하는 개념을 컴퓨터가 이해할 수 있는 형식으로 명확하고 명시적으로 정의하고 규정한 결과물'이라는 의미로 사용하고 있다. 오늘날 정보자원 관리와 관련한 온톨로지의 연구는 대략 전문용어학(Terminology) 분야에서 관심을 갖는 전문분야 온톨로지, 데이터의 의미관리를 위한 데이터 및 메타데이터 온톨로지, 시맨틱웹 분야의 웹 온톨로지, 인공지능(AI) 분야의 의미망 온톨로지(Semantic Net Ontology) 등 네 가지 갈래로 나뉘어 수행되고 있다. 메타데이터 온톨로지는 데이터와 메타데이터의 의미를 관리한다는 점에서(문헌) 정보학의 연구 영역과 가장 밀접하게 연관되는 분야이다.

데이터베이스 내에서의 중복을 방지하고 메타데이터의 등록과 검색 및 관리를 효율적으로 수행하기 위해서는 하나의 메타데이터가 데이터베이스 내에서 유일한 이름을 가져야 한다. 이를 위해서는 데이터베이스 내에서의 데이터 요소 항목이 반드시 규정에 따라 명명될 수 있도록 해야 한다. 국제표준화기구의 공동기술위원회(ISO/IEC JTC1)에서 데이터 교환 및

관리 표준화를 담당하는 분과위원회 SC32는 데이터의 의미, 구문, 표현을 표준화하기 위한 프레임워크를 개발하여 국제표준 "ISO/IEC 11179 - Metadata Registry"를 제정하였다. 이 표준은 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지·관리하게 함으로써 메타데이터의 명세와 의미를 공유할 수 있게 해준다(남영광 등 2005). 그러나 이 표준은 개념적 수준에서의 메타데이터 표준화 방법론을 제시하는 것이므로 실제에 적용시키기 위해서는 데이터 요소의 명명 규칙과 같은 세부적이고 구체적인 지침을 별도로 마련하여 시스템 구축에 반영하는 것이 필요하다.

메타데이터 명명 규칙이란 메타데이터로 사용할 데이터 요소를 명명하는 방법을 정하고 이에 대한 설명을 부가하는 것을 말한다. 명명 규칙을 수립하는 목적은 메타데이터의 이름과 속성 및 의미와 그 표현의 일관성을 확보하기 위한 것이다. 따라서 메타데이터간의 의미 일관성 유지를 위한 시스템을 구현하기 위해서는 대상 분야의 정보자원에 대한 핵심 객체 클래스를 정의하고 이들 간의 관계 및 객체 클래스의 속성(property)과 표현(representation) 용어들에 대한 분석이 선행되어야 한다.

이와 관련된 연구로서 ISO/IEC 11179 표준에 근거하여 메타데이터의 의미 공유를 위한 상세 규칙에 관한 연구가 정(Jeong 2005) 등에 의해서 수행된 바 있다. 또한 상거래와 관련하여 판매자(vender)에 관한 메타데이터를 체계화한 모형(charleston model)의 적용 연구가 패럴 등(Farrell et al, 2004)에 의해 이루어지기도 하였다. 그렇지만 이들 연구에서는 메타데이터의 의미 일관성 유지를 위한 구체적인

객체 클래스의 정의 및 이들의 속성과 표현 관계를 설명하는 작업 즉 온톨로지의 적용까지를 다루지는 않았다.

본 연구에서는 정보자원의 메타데이터를 작성할 때 메타데이터 간 의미의 일관성 유지에 적합한 메타데이터 명명 방법론을 제안하고, 이를 실제 분야에 적용할 수 있도록 메타데이터 명명 규칙의 실험적 모형을 제시하고자 한다. 이를 위해 우선적으로 ISO/IEC 11179에서 제시하고 있는 메타데이터 레지스트리 메타모형과 데이터의 기본 속성 및 개념을 논의하였으며, 이러한 논의를 토대로 객체용어(object term)와 속성용어(property term) 및 표현(presentation)에 관한 명명 규칙의 실제 적용 사례를 제시하였다. 객체용어의 생성은 자료유형의 엔터티-관계(E-R) 모형에 근거한 휴리스틱 분석을 통해 이루어졌으며, 속성용어의 명명은 더블린코어의 메타데이터 셋을 기반으로, 표현은 SHOE 1.0 버전을 기반으로 하였다.

## 2. 메타데이터 레지스트리와 메타데이터의 의미 관리

### 2. 1 메타데이터의 상호운영성과 메타데이터 레지스트리 표준화

메타데이터의 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 MARC나 더블린코어(Dublin Core)와 같은 하나의 표준적인 메타데이터로 통합하여 표현하는 방법, 자원의 속성을 감안하여 다양한 메타데이터 형식과 기술 구조를 인정하고 RDF 등을 이용하여 상호 매핑을 통해 해결하는 방법, 그리고 메타데이터 온톨로지 즉 메타데이터 레지스트리(Metadata Registry)에 의한 해결 방법의 세 가지를 들 수 있다(고영만 2005).

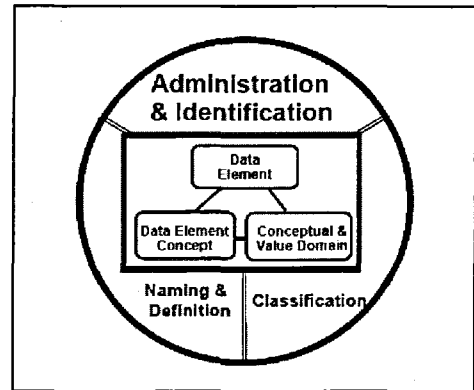
메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 관리하며, 메타데이터의 명세와 의미의 공유 즉 '공유되는 개념의 정형적 명시적 명세화'를 통해 메타데이터 셋 또는 메타데이터 요소 간의 호환성을 유지시킨다. 구체적으로 표현하면 서로 다른 데이터베이스가 같은 개념에 대해 서로 다른 식별자(태그) 혹은 서로 다른 단어를 사용할 경우 이를 해결해 주기 위해 공유되는 개념화를 정형적, 명시적으로 명세화한 결과물인 온톨로지가 집합되어 있는 메타데이터 관리시스템이라 할 수 있다. 예를 들어 표제와 제목이라는 상이한 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 상호운용성을 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. 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.

## 2. 2 메타데이터의 등록과 관리를 위한 메타모델

메타데이터 레지스트리에 메타데이터를 등록하고 관리하기 위해서는 그 형식과 방법을 포괄적으로 명세화한 모형인 메타모델이 규정되어야 한다. ISO/IEC 11179 역시 메타데이

터의 등록과 관리 모형을 명세화하기 위하여 의미와 구문(모형의 요소)을 정의한 메타모델을 제시하고 있다(그림 1 참조).



〈그림 1〉 ISO/IEC 11179 메타모델 영역

〈그림 1〉에서 보는 바와 같이 ISO/IEC 11179에서 제시하고 있는 메타모델은 관리와 식별(administration and identification), 명명과 정의(naming and definition), 분류(classification) 및 관리 항목의 데이터 요소(data element), 관리 항목의 데이터 요소 개념(data element concept), 관리 항목의 개념 영역과 값 영역(conceptual and value domain)으로 구성되어 있다(표 1 참조).

〈표 1〉 메타모델의 영역과 취급 대상

| 영역     | 취급 대상                                                                     |
|--------|---------------------------------------------------------------------------|
| 관리와 식별 | 데이터 요소 관리를 위한 등록자 정보, 제안자 정보, 담당자 정보, 관리자 정보, 등록과 생성 일자, 유효 일자, 참조 문서 등   |
| 명명과 정의 | 관리 항목의 주제 분야와 분류, 등록 대상 및 범위, 사용언어, 동의어 등의 다양한 표현방법 지원, 언어별 다양한 의미의 사용 지원 |
| 분류     | 분류체계의 유형, 분류체계 항목, 항목 간의 관계                                               |
| 관리 항목  | 데이터 요소                                                                    |
|        | 데이터 요소 개념                                                                 |
|        | 개념 영역과 값 영역                                                               |

## 2. 3 데이터 요소, 데이터 요소 개념 및 개념과 값 영역

### 2. 3. 1 데이터 요소와 데이터 요소 개념

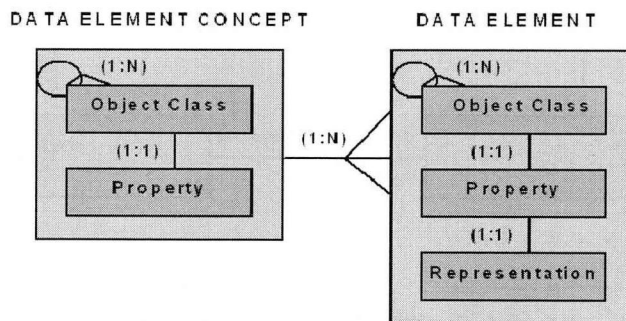
메타데이터 레지스트리에 의해 관리되는 데이터의 기본단위는 데이터 요소(data element)이며, 데이터요소는 데이터 요소 개념(data element concept)이 구체화된 것이다. 예를 들어 데이터 요소 개념에서의 국가라는 것은 국가라는 것을 생각할 때 떠오르는 개념적 표상(表象)이나 심상(心想, 예 : 통치조직을 가지고 일정한 영토에 정주하는 다수인으로 이루어진 그 무엇에 대한 심상)을 말하는 것으로 속성을 가질 수는 있으나 표현과는 아직 관련을 맺지 못하는 차원에 머물러 있으며, 데이터 요소 개념 영역에서 데이터 요소로 구체화 되었을 때 비로소 구체적인 속성과 자신의 표현을 가질 수 있게 되는 것이다(그림 2 참조).

〈그림 2〉에서 보는 바와 같이 데이터요소는 객체클래스(Object Class), 속성(Property), 표현(Representation)의 셋으로 구성된다. 객체 클래스는 메타데이터를 갖게 되는 대상을 말하며, 여기에는 명확한 범위와 의미 그리고

동일한 규칙을 따르는 속성과 행위를 가짐으로써 식별되는 실세계의 생각, 추상 또는 사물이 해당된다. 객체클래스는 외연적 정의와 내포적 정의를 통해 규정되며, 전통적으로 개념의 내포(Intension)에 해당하는 것을 전문용어학에서는 특질(Characteristics/Chracteristika/Merkmale), 전산학과 지식공학에서는 속성(Attribute, Porperties/Eigenschaften), 등으로 불려왔다. 표현은 데이터를 표현하기 위한 기호화/상징화에 관한 것으로서, 데이터의 표현은 데이터의 값 영역(value domain)과 데이터 유형(data type) 등에 의해 좌우 된다.

### 2. 3. 2 개념 영역과 값 영역

개념 영역(conceptual domain)은 값이 가지는 의미의 집합체로서 값 영역(value domain)의 집합체의 포괄적 표상 또는 심상에 속하는 것이다. 개념 영역은 데이터 요소의 형태로 표현되는 것으로서 데이터 요소 개념의 대상, 예를 들면 국가명의 대상이 되는 '(전 세계의) 국가'가 이 영역에 해당된다. 개념 영역의 국가는 국가라는 말을 들었을 때 떠오르는 개념적 심상인 '통치 조직을 가지고 일정한 영토에 정



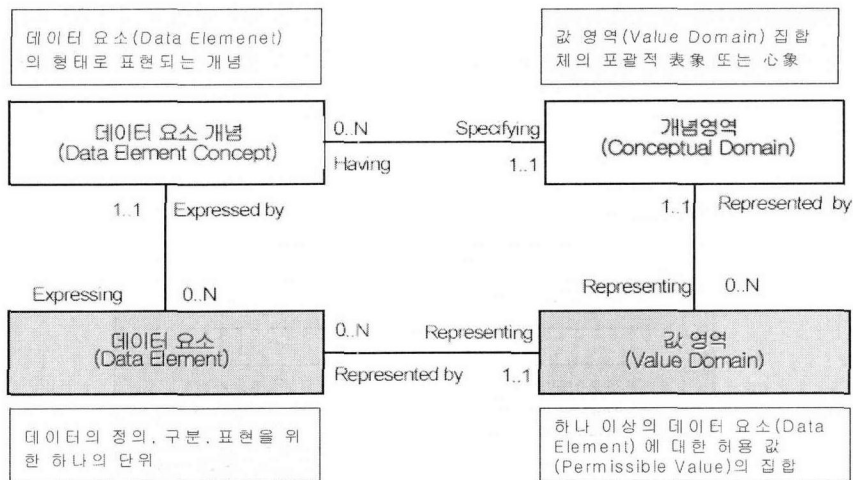
〈그림 2〉 데이터 요소 개념과 데이터 요소의 관계

주하는 다수인으로 이루어진 그 어떤 대상'을 의미한다. 개념 영역에 속하는 눈에 보이지 않는 심상 차원의 국가들은 허용 값(permissible value, 예: 국가명)을 통해 사람에게 인지시킬 수 있으므로 허용 값이 갖는 의미 즉 값 의미(value meaning)의 집합이라고도 할 수 있다.

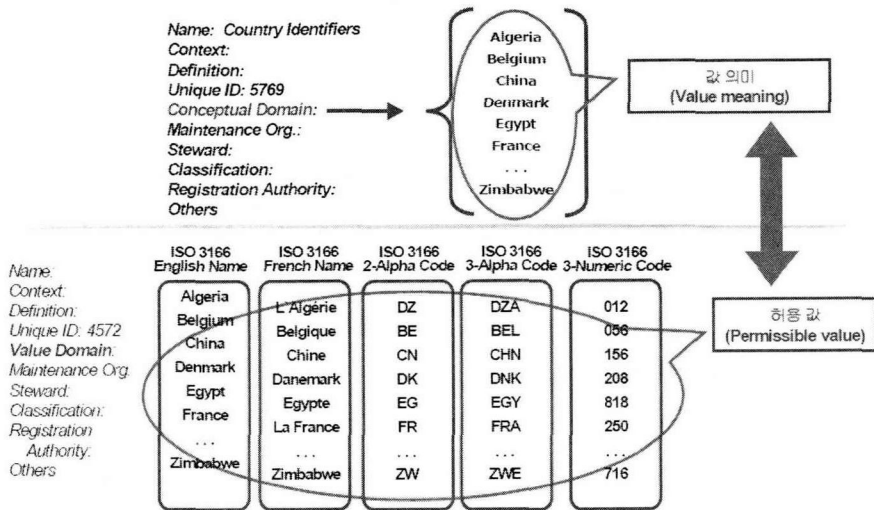
이때 값 영역은 영문 국가명으로 표현되는 하나하나의 개별 값 즉 Korea, France, England, Canada 등과 같은 유효한 값들의 영역이고 이 값 영역의 대상이 되는 값의 의미인 '한국으로 지정된 국가', '영국으로 지정된 국가' 등의 집합이라 할 수 있는 '전 세계의 국가'가 개념 영역에 해당한다. 개념 영역 상에는 존재하지만 값으로는 표현할 수 없는 것들이 있기 때문에 개념 영역은 값 영역보다 그 범위가 크며, 모든 값 영역은 간접적으로 데이터 요소 개념을, 직접적으로는 개념 영역을 표시하게 된다. 또한 허용 값은 하나 또는 그 이상의 데이터 요소에

대하여 유효한 값을 의미하며 메타데이터 레지스트리를 만드는 목적에 따라 그 범위가 달라진다. 따라서 값 영역은 데이터 요소의 표현이나 개념 영역과 연결되어 있어 데이터 요소와는 별개로 개념을 표시한다(그림 3 참조).

〈그림 4〉는 'ISO 3166 - Country Code' 표준에 따라 국가명을 등록할 때 값 영역과 개념 영역과의 관계를 보여준다. 우리나라를 의미하는 개념 영역의 눈에 보이지 않는 '한국'이라는 심상의 차원이 있으며, '대한민국', 'Korea', 'ROK', 혹은 '한국' 등은 개념 영역의 '한국'이라는 심상을 표현하는 값이다. 하나의 개념을 여러 가지 방법으로 표현할 수 있기 때문에 값 영역에는 여러 가지 허용 값들이 포함될 수 있다. 〈그림 4〉는 알제리아라는 국가에 대한 'DZ', 'DAZ', '012' 등의 허용 값을 보여준다(한국데이터베이스진흥센터 2005, 10).



〈그림 3〉 개념 영역, 값 영역, 데이터 요소 개념, 데이터 요소의 관계



〈그림 4〉 개념 영역과 값 영역의 관계 사례

### 3. 메타데이터 온톨로지와 메타데이터 명명 규칙

#### 3. 1 메타데이터 온톨로지

온톨로지는 개념(concept)과 개념화(conceptualization)의 두 가지 인식을 기반으로 이루어진다. 개념화는 사물이나 표상의 성질, 공통성, 본질을 추출하고 파악하는 과정을 말한다. 개념은 사람들이 머리 속에 가지고 있는 일반적, 추상적인 지식, 관념, 생각으로 개념화 과정의 결과로 생성되며 언어를 포함한 기호로 표현된다(옥철영 2004). 사람들 간의 의사소통이 원활히 이루어지기 위해서는 동일한 대상에 대하여 공유할 수 있는 기의가 형성되어야 하는 것과 마찬가지로 전산 시스템 간에 의미 전달이 이루어지기 위해서는 기표에 해당하는 데이터 요소와 값 영역의 구조 및 의미를 공유

할 수 있어야 한다. 이를 위해서 필요한 것이 바로 데이터 요소의 온톨로지라 할 수 있으며, 따라서 메타데이터 온톨로지란 대상과 속성 및 표현으로 구성되는 메타데이터의 요소가 가지는 공유하는 명시적 개념화의 결과물이라 할 수 있다.

임의의 대상에 관한 개념과 개념화 및 개념의 기호화의 관계는 기호학을 이해하기 위한 기본 모형인 의미 삼각도를 통해 파악할 수 있다. 예를 들어 사람들은 이 세상에 존재하는 = 나무=라는 대상의 개념 즉 심상으로서의 「나무」를 마음 속에 가지고 있으며, 대상으로서의 나무나 개념으로서의 나무를 “나무”, “목”, “Tree” 등의 기호로 표현한다. 그리고 “나무”라는 말을 듣거나 기호(글자)를 볼 경우 「나무」라는 개념을 마음 속에 그릴 수 있게 된다(그림 5 참조).

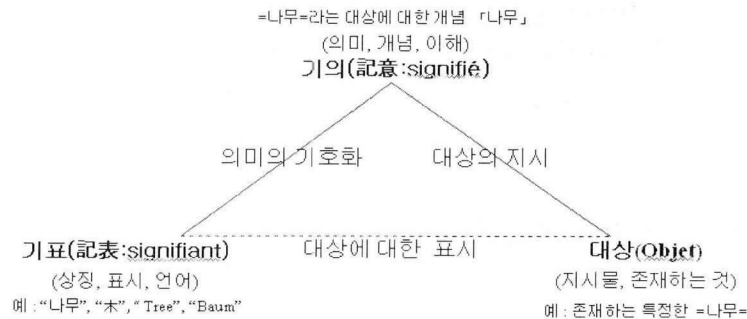
데이터 요소와 데이터 요소 개념 및 개념 영

역을 의미 삼각도를 통해 살펴보면, 데이터 요소는 대상과 대상의 속성 및 그 속성의 표현인 기표로 이루어지며 데이터 요소 개념은 표현 즉 기표와는 무관한 데이터 요소들의 개념적 심상인 기의에 해당한다. ISO/IEC 11179에서는 기의에 해당하는 것으로서 데이터 요소 개념(Data Element Concept)과 개념 영역(Conceptual Domain)을 사용하며, 기표에 해당하는 것으로서는 데이터 요소(Data Element)와

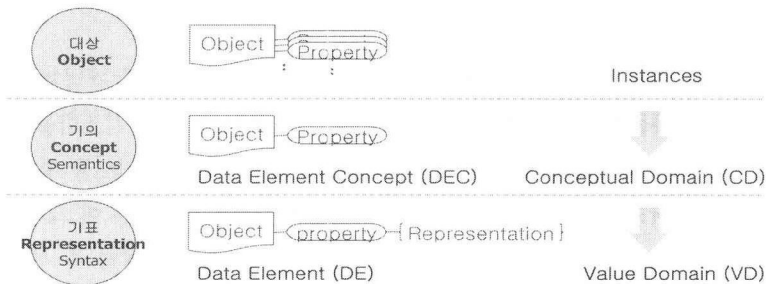
값 영역(Value Domain)을 사용하고 있다(그림 6 참조).

### 3. 2 메타데이터 명명규칙

하나의 정보 객체를 표현하기 위해서는 그 객체에 관련된 속성들과 속성 값의 표현 방법이 정해져야 한다. ISO/IEC 11179-1에서는 <표 2>와 같은 방법으로 메타데이터 프레임워크를



〈그림 5〉 기의, 기표, 대상의 의미 삼각도



〈그림 6〉 기의, 기표, 대상과 메타데이터 온톨로지 모형

〈표 2〉 ISO/IEC 11179 Part 5의 메타데이터 명명 프레임워크

| 수준 | 항목        | 항목의 의미         | 사례                 |
|----|-----------|----------------|--------------------|
| 개념 | 데이터 요소 개념 | 객체클래스 + 속성     | 국가 이름              |
|    | 개념 영역     | 값 의미의 집합       | 이 세상의 국가들(예: 대한민국) |
| 표현 | 데이터 요소    | 데이터 요소 개념 + 표현 | 국가 이름 코드           |
|    | 값 영역      | 허용 값의 집합       | ISO 639-2(예: KOR)  |

구성하고 있다. 하나의 데이터 요소(DE: Data Element)는 객체의 속성들의 결합으로 이루어진 데이터 요소 개념(DEC: Data Element Concept)에 표현 부분이 더해짐으로써 얻어진다. 그리고 하나의 데이터 요소에 상응하는 값은 그 데이터 요소에 상응하는 값 의미(VM: Value Meaning)를 표현할 수 있도록 규정한 허용 값(PV: Permissible Value)의 집합 중에서 부여되게 된다.

ISO/IEC 11179 Part 5에 따르면 데이터 요소의 명명을 위한 일반 규칙은 크게 의미, 구문구조, 어휘에 관한 규칙으로 세분된다. 의미에 관한 규칙은 데이터 요소 항목의 이름 각 부분이 뜻하는 바와 각 부분을 분리해 주는 구분자에 관한 규칙을 말한다. 구문 구조에 따라 도출된 데이터 요소 항목명의 각 부분을 통해 다른 데이터 요소 항목들과의 관계나 분류 체계를 알 수 있다. 구문 구조에 관한 규칙은 이름을 구성하는 각 단어들의 배열에 관한 규칙이다. 구문 구조는 절대 배열과 상대 배열 등이 있다. 한정어처럼 위치가 고정되어 있지 않고 수식하는 단어의 바로 앞에 위치시키는 규정을 상대 배열, 속성처럼 반드시 객체 클래스 다음에 위치시키는 규정을 절대 배열이라고 한다. 어휘에 관한 규칙은 데이터 요소 항목명에 사용할 어휘의 선택, 선호 용어, 비선호 용어, 동의어, 축약어, 문자의 길이, 철자법, 허용 가능한 문자 등에 관한 규칙으로 명명의 일관성을 위해 모든 데이터 요소 항목에 적용해야 한다.

이와 같이 ISO/IEC 11179의 Part 5에서 제시하는 프레임워크를 바탕으로 메타데이터 명명 지침을 구성할 경우 하나의 데이터 요소

에 대한 명명 규칙은 최소한 기본 규칙과 한정어 부가 규칙으로 나뉘어 진다 :

- 기본 규칙: [객체 용어] + [속성 용어] + [표현 용어]
- 한정어 부가 규칙: [(한정어)+객체 용어] + [(한정어)+속성 용어] + [(한정어)+표현 용어]

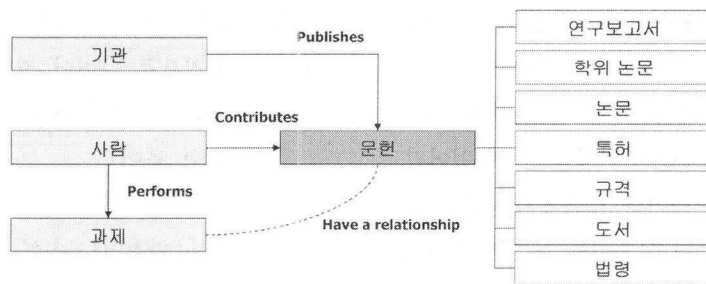
## 4. 데이터 요소 명명 규칙의 적용 사례

### 4. 1 데이터 요소의 객체 온톨로지 적용

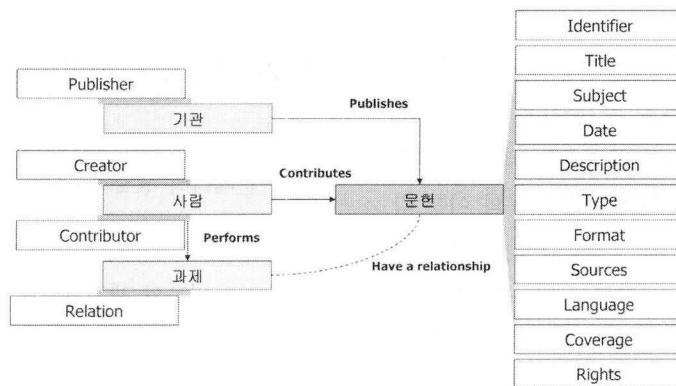
학술정보의 경우 그 생성과정을 살펴보면 일반적으로 정보의 생성과정은 사람이 과제를 수행함으로써 특정 문헌의 생성에 기여를 하게 되며 과제의 발주 기관을 통해 해당 문헌이 발행된다(그림 7 참조). 이 과정에서 최소한 문헌, 사람, 기관, 과제 등 4가지의 객체가 각각의 기능적 위치를 갖게 되며, 문헌의 하위 구분으로는 연구보고서, 학위논문, 특허, 규격, 도서, 법령 등을 들 수 있다. <그림 7>은 학술정보의 생성에서 나타나는 객체 측면의 용어를 규정하고 이 용어들 간에 존재하는 관계의 집합을 도식화한 것으로서 학술정보 생성의 객체 측면에서 온톨로지를 구축할 수 있도록 개념화한 것이다(Kim 2005).

### 4. 2 데이터 요소의 속성 온톨로지 적용

전산 시스템에서 문헌이라는 객체를 설명하기 위해서는 여러 가지 속성을 활용하게 된다. 더블린 코어 15개 요소를 학술정보의 생성과정



〈그림 7〉 학술정보 생성과정 및 객체 온톨로지 적용 대상



〈그림 8〉 학술정보의 객체와 더블린 코어 요소의 적용

에서 나타나는 객체의 속성으로 활용할 경우 문헌 기술을 위한 11개의 속성과 문헌 이외의 객체(사람, 기관, 과제)를 기술하는 4개의 속성으로 구성된다(그림 8 참조).

더블린 코어의 15개 기본 요소는 일반적인 개념 수준이므로 적용되는 도메인에 따라 그 명칭을 변형하여 사용할 수 있다. 예를 들면 Title에 해당하는 요소가 논문에 적용될 경우에는 우리말로 '제목'이 되지만 특허에 적용될 경우에는 발명의 '명칭'으로 하는 것이 적절하다. 더블린 코어의 이러한 특성을 고려하여 학술정보의 대표적인 7개 문헌 유형에 대하여 공유할 수 있는 속성 용어를 정리하면 〈표 3〉과 같다. 이 용어들은 궁극적으로 데이터 요소

의 이름을 정할 때 속성 부분에 사용될 것이며, 〈표 3〉에서 제시되는 속성 용어는 최소한의 용어이므로 필요한 경우 추가하여 활용할 수 있다.

#### 4. 3 데이터 요소의 표현 온톨로지 적용

데이터 요소에서의 표현은 속성을 구체화한 것으로서 데이터 요소의 값으로 어떤 내용이 될 것인가를 알 수 있게 해주는 부분이다. SHOE 1.0 버전은 표현 용어의 데이터 유형(data type)은 String(번호, 유형, 주소, 설명, 이름 등), Number(수, 액 등), Date(년, 일자 등), Truth(여부)로 나누고 있다. 속성 용어는

&lt;표 3&gt; 학술정보의 유형별 속성 및 속성 온톨로지 적용 대상

| 문헌유형<br>DC요소 | 연구<br>보고서  | 학위<br>논문 | 논문      | 특허                   | 규격         | 도서            | 법령   |
|--------------|------------|----------|---------|----------------------|------------|---------------|------|
| Identifier   | 번호         | 학번       |         | 번호                   | 번호         | ISBN          | 번호   |
| Title        | 명          | 제목       | 제목      | 명칭                   | 명          | 제목            | 이름   |
| Subject      | 분류, 색인어    | 색인어      | 분류, 색인어 | 분류, 색인어              | 분류         | 분류            |      |
| Date         | 발행일        | 취득일      | 발간년도    | 출원일, 공개일<br>공고일, 등록일 | 제정일, 개정일   | 발행일           | 공포일  |
| Description  | 초록, 목차     | 초록, 목차   | 초록      | 초록, 청구범위             | 초록, 목차     | 초록, 목차,<br>가격 | 원문   |
| Type         | 유형         | 구분       | 유형      | 문서유형                 | 문서유형, 부문   | 유형            | 유형   |
| Format       |            |          |         |                      |            |               |      |
| Sources      | 원문         | 원문       | 원문      | 원문                   | 원문         | 원문            | 원문   |
| Language     | 언어         | 언어       | 언어      | 언어                   | 언어         | 언어            |      |
| Coverage     |            |          |         | 지정국                  |            |               |      |
| Rights       | 공개         |          |         | 심사청구<br>심사청구일        |            |               |      |
| Creator      | 책임자, 연구원   | 저자       | 저자      | 발명자, 권리자             |            | 저자            |      |
| Contributor  |            | 지도교수     |         | 대리인                  | 담당자        | 편저자           |      |
| Publisher    | 발행기관       | 학교       | 발행기관    |                      |            | 출판사           |      |
| Relation     | 관련과제, 참고문헌 | 참고문헌     | 참고문헌    | 우선권                  | 관련규격, 인용규격 | 참고문헌          | 관련법령 |

&lt;표 4&gt; 학술정보의 표현 용어 및 표현 온톨로지 적용 대상

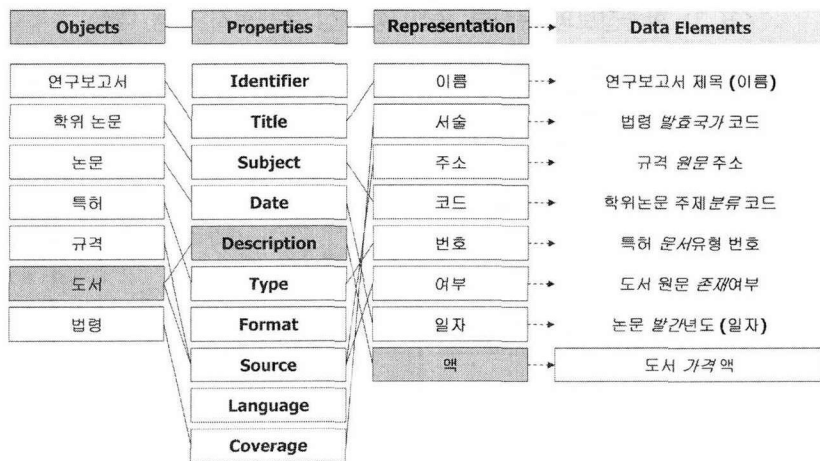
| 구분   | 고정 길이                           | 가변 길이                                                        |
|------|---------------------------------|--------------------------------------------------------------|
| 열거형  | 코드(code), 번호(number), 여부(truth) | 유형(type)                                                     |
| 비열거형 | 년(year), 일자(date)               | 수(number), 액(amount), 주소(address), 설명(description), 이름(name) |

또한 해당 값의 길이가 고정되어 있는 경우와 고정되어 있지 않은 경우로 나누어지며, 열거형과 비열거형으로 나눌 수 있다. 국내에서 제정되었거나 만들어지고 있는 산업정보 메타데이터 사례에서의 표현 용어는 <표 4>와 같으며, 도메인이 다른 영역으로 바뀌게 될 경우 수정되거나 추가될 수 있을 것이다.

#### 4. 4 온톨로지 기반 메타데이터 명명 규칙 적용 사례

지금까지 살펴본 학술정보의 객체와 유형별

속성 및 표현을 순차적으로 조합할 경우 <그림 9>와 같은 데이터 요소를 추출함으로써 온톨로지를 적용한 명명규칙을 구성할 있게 된다. 물론 조합 과정에서 속성과 표현 용어는 일정한 기준에 따라서 변형될 수 있다. 예를 들면, '연구 보고서 제목'의 경우 '연구보고서 제목 이름'으로 하여야 하지만 '제목'과 '이름'이 중복성이 있으므로 이 중 하나를 생략해서 표현할 수 있다. 또 다른 경우로는 '도서 가격 액'의 경우 '도서 가격서술 액'이라고 하여야 하지만 '서술'을 생략하는 것이 호칭 상 효율적이다.



〈그림 9〉 온톨로지 기반의 데이터 요소의 명명 규칙 적용

## 5. 결 론

본 연구에서는 ISO/IEC 11179에서 제시한 데이터 요소와 데이터 요소 개념 및 이들의 상위 메타 모델에 관하여 논하였으며, 이를 토대로 지식정보의 핵심을 이루는 문헌정보의 메타데이터를 효과적으로 표준화하기 위한 명명 원칙과 방법론을 ISO/IEC 11179 표준과 더블린 코어에 입각하여 제시하였다. 방법론의 제시에 있어서 본 연구는 특히 메타데이터 관련 용어와 이들 용어 간의 관계를 명확하게 분석한 온톨로지를 구성함으로써 메타데이터 명명의 구조적, 의미적 일관성을 유지할 수 있도록 하였다.

본 연구에서 제시된 객체와 속성 및 표현의 조합을 통한 데이터요소의 추출 및 온톨로지 적용은 도메인별 메타데이터의 세부적 명명 규칙을 구성하는 데 있어서 하나의 방법론으로 사용될 수 있을 것이며, 전문가들에 의해서 체계적

으로 정리되어 사실 표준으로 인정될 경우 실제에 적용하는 데 무리가 없을 것이다. 제시된 메타데이터 명명 방법론은 국내에서 제정되는 지식정보 분야의 메타데이터 표준 간에 발생하고 있는 일관성 부재로 인한 혼란을 최소화하고 메타데이터 표준화 관련 지식의 성숙과 관련 기술 수준 제고에도 기여할 수 있을 것이다.

다만 본 연구는 수작업에 의한 온톨로지를 지향한 방법론을 제시한 것으로서 향후 표준화 절차에 관한 연구 및 학술정보 이외의 분야에서 다양한 적용과 검증 연구가 수행되어야 할 것이다. 보다 복잡하고 다양한 속성을 지닌 객체의 경우 전산적인 도구를 개발하는 연구로 발전되어야 할 것이나, 전산적인 온톨로지 도구의 개발이 수작업에 의한 온톨로지를 대체하기에는 더 많은 요소 기술들의 개발과 지원이 병행되어야 함을 인식하는 것이 매우 중요하다.

## 참 고 문 헌

- 고영만. 2005. 온톨로지와 웹 온톨로지. 『메타데이터 표준화 포럼 제1회 워크숍 자료집 - 메타데이터와 온톨로지』. 2005.3.9, 서울: 한국과학기술정보연구원.
- 남영광, 서태설, 황상원. 2005, ISO/IEC 11179 표준에 따른 산업기술정보 메타데이터 표준화. 『정보관리연구』. 36(1): 57-75.
- 옥철영. 2004. 한국어 정보처리와 온톨로지. 『2004 한국어정보처리연구회 동계 튜토리얼』. 2004.2.13, 서울: 성신여대.
- 한국데이터베이스진흥센터. 2005, 메타데이터 레지스트리 구축 지침서. 2005.10.
- Farrell, K. T. and M. Truitt 2004, Defining Functional requirements for Acquisitions Records: Vendor Metadata, *Library Collections, & technical Services*, 28: 473-487.
- Jeong, D., P. H. In, F. Jarnjak, Y.-G. Kim, and D.-K. Baik 2005. A Message Conversion System, XML-based Metadata Semantics Description Language and Metadata Repository. *Journal of Information Science*, 31(5): 394-406.
- Kim, H. H. 2005. ONTOWEB: Implementing an Ontology-Based Web Retrieval System. *JASIST*, 56(11): 1167-1176.
- ISO/IEC 11179 Information technology -- Metadata registries(MDR) Part 1 : Framework.
- ISO/IEC 11179 Information technology -- Metadata registries(MDR) Part 3 : Registry Metamodel and Basic Attributes.
- ISO/IEC 11179 Information technology -- Metadata registries(MDR) Part 5 : Naming and Identification Principles
- ISO 15836:2003: Information and documentation-The Dublin Core metadata element set.
- KSXISOIEC11179-5, 정보기술 - 데이터 요소의 명세와 표준화 - 제 5부 : 데이터 요소에 대한 명명 및 식별원칙.
- TTAS.IS-11179.1 정보기술 - 데이터 요소의 명세와 표준화 - 제1부 : 데이터 요소의 명세와 표준화를 위한 기본틀.
- 국가메타데이터레지스트리. [cited 2005.11.10] <<http://www.metadata.go.kr>>
- SHOE 1.0 [cited 2005.11.10] <<http://www.cs.umd.edu/projects/plus/SHOE/spec1.0.html>>