

INDECS와 기존 메타데이터간의 상호운용성에 관한 연구*

A Study on the Interoperability of INDECS and Metadata

윤세진, 오경목 ; 숙명여자대학교 문헌정보학전공
황상규 ; 한국국방연구원

Se-Jin Yun, Kyung-Mook Oh ; Sookmyung Womens University
Sang-kyu Hwang ; KIDA

본 연구에서는 현재 DOI를 위한 메타데이터로 채택되어 활발하게 연구되고 있는 전자상거래를 위한 메타데이터 INDECS(Interoperability of Data in E-Commerce Systems)를 전통적 형태의 메타데이터인 MARC21 (Machine Readable Cataloging)과 더블린코어(Dublin Core) 메타데이터간의 데이터 요소들을 비교하여, 서로 다른 메타데이터간의 상호운용성을 제공하기 위한 방안에 대하여 연구하였다.

1. 서론

메타데이터란 문헌의 속성을 기술하기 위한 데이터(레코드의 이름, 설명, 크기 혹은 길이, 구성요소로써 데이터 항목들 그리고 데이터 집합과 주요키와 보조키에 대한 식별 등에 관한 정보)를 말한다. 이러한 메타데이터는 어떤 목적과 용도에 따라 콘텐츠마다 다양하게 주어지기 때문에 상이한 메타데이터간의 호환성을 제공하기 위한 상호운용성이 필요하다.

상이한 메타데이터 스키마가 상호호환되고 특정 분야에서 개발된 메타데이터가 다른 분야에서도 이용 가능하도록 하는 것을 목표로 만들어진 것이 바로 INDECS 메타데이터이다.

INDECS는 1998년 유럽연합이 지원하는 Info2000 프로젝트의 지원을 받아, 국제적인 저작권 소유자들의 주도로 시작된 프로젝트이다.

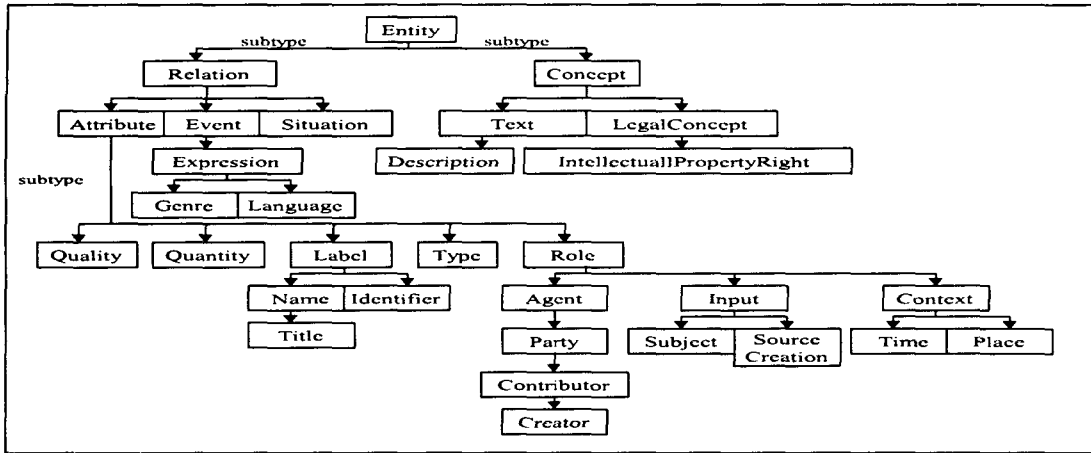
프로젝트 초기모델은 더블린코어를 기초로 발전하였지만 지금은 새로운 형태를 띠고 있다.

2. INDECS의 구성요소

INDECS의 구성요소는 <그림 1>에서 볼 수 있듯이 크게 Entity, Attribute, Relation으로 구분할 수 있다. INDECS에서 Entity는 식별되어지는 어떤 것(something)으로 World Wide Web Consortium에 의해 채택된 resource의 의미와 비슷하다. INDECS에서는 대상이 되는 Entity를 3가지 관점인 The general view, The commerce view, The intellectual property view로 볼 수 있다. <그림 1>은 The general view에서 본 INDECS의 기본적인 구성요소들이다.

INDECS모델의 모든 Entity는 5가지 속성 (labels, quantities, qualities, types, roles)을 갖

* 이 연구는 과학기술부 지원 '01 숙명여자대학교 연구기반확충사업 (01-N6-02-01-A-01)에 의하여 지원되었음.



<그림 1> INDECS의 기본적인 구성요소

는다. 다섯 가지 속성 각각은 그것만의 특별한 구조와 행동을 가지고 있으며 다양한 종류의 Entity를 위한 상호운용의 일반적인 구조를 제공한다.

Relation은 INDECS모델에서 가장 중요한 구조로 Event, Situation으로 구성된다. Event는 두 개 이상의 Entity간의 동적인 관계(dynamic relation)에서 발생되어지는 사건들을 가리킨다. INDECS에서는 모든 저작물(work)의 생성, 유통, 이용을 포함하여 지적재산권의 획득, 실현, 이전, 이용까지 Event를 중심으로 표현하고 있다. Situation은 Entity간의 정적인 관계(static relation)로서 '소유(possessing situation)'와 같은 시간이 흘러감에 따라 쉽게 변하지 않는 관계를 의미한다. 이러한 Relation은 둘 또는 그 이상의 Entity들의 관계에서 어떠한 역할(role)을 수행하게 된다.

역할(role)은 다시 agent(주체), input(투입물), output(결과물), context(상황)로 구성된다.

이러한 INDECS는 유일한 식별성(unique identification)과 기능적 세분성(functional granularity)을 기본 원칙으로 하고 있으며, 그 속성상 콘텐츠의 고정된(static) 데이터요소 보다는 동적인(dynamic) 데이터 표현에 중점을 두고 있다.

지금까지 살펴보았듯이 INDECS모델에서는 대상 자원에 대한 메타데이터 체계 구성에 있어, 기존 메타데이터 방식과는 상이한 독자적인 개념체계를 가지고 있다. 따라서, 메타데이터 작성자가 사전에 INDECS모델에 대하여 보다 명확한 개념적 이해를 하지 못할 경우, 원활한 작업 수행이 이루어질 수 없다.

3. MARC21, 더블린코어, INDECS 데이터 요소 비교

다음 <표 1>은 INDECS의 대표적인 데이터 요소와 기존의 메타데이터인 더블린코어, MARC21의 데이터요소를 비교한 것이다. 더블린코어가 15개의 단순하고 명료한 데이터요소로 구성되었기 때문에 <표 1>에서 기준으로 하였다.

INDECS의 메타데이터 구성요소에서 사용한 기호들에 대해 간략히 설명하면 '_'(underscore)는 처음에 오는 용어가 마지막으로 오는 용어를 형식적으로 제한한 것으로 'Title (Creation_label)'은 INDECS의 'Title'이란 데이터 요소가 "창작물(Creation)에 부여되는 라벨(labe)"이라고 그 의미를 한정짓는다. 또 다른 기호인 '/'(forward slash)는 subtype관계를

<표 1> DC, INDECS, MARC21 data element의 비교

DC data element	INDECS data element	MARC data element
1 Title	Title(creation_label)	245 00\$a, 246 33\$a, 250 ##\$a, 242 00\$a, 130 0#\$a, 247 10\$a
2 Creator	Creator((original_creation). contributor)	100 0#\$a, 700 1#\$a, 710 2#\$a, 711 2#\$a, 720 ##\$a, 720 ##\$e
3 Subject	Subject(input/)	050 ##\$a, 082 ##\$a, 600 17\$a, 610 27\$a, 611 27\$a, 650 #0\$a, 650 #7\$a, 653 ##\$a, 072 #7\$a.
4 Description	Description(text/)	500 ##\$a, 505 0#\$a, 520 ##\$a, 520 ##\$b, 521 ##\$a, 586 ##\$a
5 Publisher	Contributor(party/)	260##\$a, 260##\$b
6 Contributor	Contributor(party/)	100 0#\$a, 700 1#\$a, 710 2#\$a, 711 2#\$a, 720 ##\$a, 720 ##\$e
7 Date	Time(context/)	260 ##\$c, 260 ##\$g, 307 ##\$a, 518 ##\$a, 541 ##\$d
8 Type	Genre(expression.form/)	516 ##\$a, 655 #7\$a
9 Format	Quantity(attribute/)	300 ##\$a, 300 ##\$c, 340 ##\$a, 856 ##\$q, 007, 008
10 Identifier	Identifier(unique_label)	020 ##\$a, 022 ##\$a, 024 8#\$a, 028 01\$a, 856 40\$u, 856 #0\$g
11 Source	SourceCreation(input/)	786 0#\$n, 786 0#\$o
12 Relation	Situation(static_relation/)	787 0#\$n, 787 0#\$o, 776 0#\$n, 776 0#\$o, 773 0#\$n, 773 0#\$o, 774 0#\$n, 774 0#\$o, 775 0#\$n, 775 0#\$o, 786 0#\$n, 786 0#\$o, 510 0#\$a, 538 ##\$a, 490 1#\$a, 490 1#\$x, 780 00\$t, 785 00\$t
13 Language	Language(verbal_expression.form/)	041\$a, 546 ##\$a,
14 Coverage	Place, Time(context/)	522 ##\$a, 513 ##\$b
15 Right	IPR(legalConcept)	540 ##\$a, 540 ##\$b,

의미하며, 'Time(context/)'은 Role의 Agent, Input, Context 중에서 마지막 Context의 subtype임을 의미한다. 마지막으로 '.'(full stop)은 두 엔티티간의 attribute관계를 표시하며, 'expression.form'은 'expression'의 attribute임을 의미한다

더블린코어와 INDECS의 각 데이터 요소에 대해 살펴보면 <표 1>에서 볼 수 있듯이 더블린코어의 Title, Creator, Subject, Description, Identifier, Language는 INDECS의 subtype에서 같은 용어로 그대로 사용하고 있다. 용어는 다르지만 같은 의미로 사용되는 데이터 요소로는 Date(Time), Format(Quantity), Source(Source-

Creation), Coverage(Place, Time)이다. 더블린코어의 Contributor와 Publisher는 INDECS에서 Contributor로 묶어서 producer(제작자), operator(조작자), recorder(등록자) 등으로 세분화하여, 단순히 책을 출판하는 출판사의 개념에서 벗어나 다양한 콘텐츠를 포용할 수 있는 범위로 확장되어 표현되고 있다.

더블린코어의 Type은 INDECS에서 멀티미디어(multimedia), 뮤지컬(musical), 화보(pictorial) 등의 subtype을 포함하는 Genre에 해당된다, 그러나 INDECS에도 Type이 있는데 INDECS에서의 Type은 'Entity그룹에 속하는 Entity를 하나 또는 그이상의 특징들로 분류한

것'이라고 정의하고 있으며, 더블린코어의 Type과는 그 용도가 서로 다르다. 단지 최신버전의 DOI핸드북 후반부에서 INDECS에서의 Type은 더블린코어의 Type에 해당하는 Genre 뿐만 아니라 Audience(이용자), Format(형식), IType(저작권 유형) 등을 포함하는 것으로 더블린코어에서 보다 더 포괄적인 의미로 사용된다는 것을 제시해 주고 있다.

더블린코어의 Relation은 INDECS의 Relation 중에서도 단순한 정적인 관계를 나타내는 Situation에 해당된다. 더블린코어에서 Relation은 단순히 자원과 자원의 관계를 나타내 주는 것이지만, INDECS의 Relation은 자원과 자원뿐만 아니라 자원과 사람, 사람과 사람 등으로 다양한 관계를 나타낼 수 있다.

이렇듯 Relation이 다양하게 나타나는 것은 INDECS가 지니는 특징 중에 하나인 다양한 관점(view)에 따라 Relation의 역할(role)이 달라지기 때문이다.

마지막으로, INDECS와 MARC21간의 메타데이터 요소들간의 비교를 위하여, INDECS메타데이터 모델을 보다 구체화하여 실제 메타데이터 품을 제시한 ONIX(Online Information Exchange)의 구성항목들을 INDECS의 프레임 구조에 맞추어 재분류하는 작업을 선행하였다. 다음 단계의 작업으로는 INDECS와 ONIX간의 매핑 결과물과 ONIX와 MARC21과 비교, 분석한 Library of Congress의 연구를 기초로, 최종적으로 INDECS와 MARC21간 서로간의 대응되는 항목들을 결정하였다.

6. 결론

앞의 <표 1>에서 살펴보았듯이 더블린코어의 거의 모든 메타데이터 요소를 INDECS는 수용하고 있다. 그러나 더블린코어는 동적인 데이터 표현에 중점을 두고있는 INDECS의 메타데이터 요소들을 포용하지 못하고 있음을 확인할 수 있다.

Library of Congress에서 만들어 놓은 ONIX와 MARC21의 매핑 테이블을 보면 MARC21의 데이터 요소들과 INDECS메타데이터 모델을 구체화시킨 ONIX 데이터 요소들이 거의 대부분 매핑되고 있다. 이는, MARC21 역시 동적인 데이터 표현에 중점을 두고있는 INDECS의 모든 요소들을 수용하기에는 현실적으로 불가능하지만, 실제 메타데이터 운영상에 있어서 MARC21과 INDECS간에는 어느 정도 서로간의 상호운용성이 보장됨을 확인할 수 있었다.

마지막으로, 각각의 더블린코어와 INDECS, MARC21의 세부항목을 비교해보았을 때, 세 가지 모두 저작권(Right)과 관련된 항목들이 아직까지는 미흡함을 확인할 수 있었다. 따라서 세 가지 메타데이터 모두, 새로운 환경변화에 맞추기 위해서는 저작권에 관련된 요소가 보장되어야 함을 확인할 수 있었으며, 앞으로 보다 다양한 메타데이터간의 상호운용성을 위한 추가연구가 지속적으로 이루어져야 할 것이다.

참고 문헌

- 이석기, 1999. 디지털 도서관 구축을 위한 메타데이터 코딩 방안. 『디지털도서관』, 1999년 겨울호, 74-90.
- DCMI. 1999. "Dublin Core Metadata Element Set, Version 1.1:" Reference Description." [cited 2001. 5. 31].
<<http://dublincore.org/documents/dces/>>
- Library of Congress, 2000. "MARC 21 Bibliographic Data" [cited 2001. 3. 13].
<<http://lcweb.loc.gov/marc/bibliographic/ecbdhome.html>>
- Rust, Godfrey and Bide, Mark. 2000. "The <indec> metadata framework" [cited 2000. 12. 9.]. <<http://www.indec.org>>