

메타데이터 통합 방안

Integration of Different Metadata Formats

심 경 (Kyung Shim)*

〈목 차〉

- | | |
|--------------------------------|--------------------------------|
| I. 서론 | IV. 메타데이터 포맷의 변환과 통합 |
| II. 메타데이터: 역할, 용어의 출현 및 개념의 혼돈 | 1. 특정 메타데이터 포맷의 기존/다른 형식으로의 변환 |
| III. 새로운 메타데이터 포맷의 발생 | 2. 범용적 구조를 통한 통합 |
| 1. 웹 자원을 위한 메타데이터 포맷의 발생 | V. RDF를 이용한 메타데이터 통합 |
| 2. 일반 자원을 위한 메타데이터 포맷의 발생 | VI. 메타데이터 레지스트리를 통한 데이터 요소 표준화 |
| | VII. 결론 |

초 록

메타데이터 개념에 관한 혼란을 살펴보고 개념 변화의 이유에 대하여 알아보았다. 이중 메타데이터 포맷에 관하여 새로운 포맷의 생성 이유, 그들의 특징, 특히 MARC 포맷과 관련하여, 그리고 이들의 변환 및 통합 방법과 문제점을 배경연구로 조사하였다. 이중 메타데이터의 통합 틀로서 RDF와 메타데이터 레지스트리를 조사하고 비교하였다. 이 두 방법은 서로 경쟁적인 접근방식이 아닌 상호보완적인 관계를 가진다고 결론 내렸다.

주제어: 메타데이터, 메타데이터 포맷, RDF, 표준화

Abstract

Semantic quirks about the concept of metadata are briefly probed. The reasons behind the conceptual drift are also reviewed. Regarding different formats of metadata, the rationale behind the creation of new formats, their characteristics in relation to MARC format, and problems and methods of conversion or integration of metadata are investigated as background study. RDF (Resource Description Framework) as an integration tool for different metadata standards and Metadata Registry as a candidate for the purpose are studied and the both are compared. It is concluded that two approaches are not mutually exclusive, rather they are complementary.

Key Words: metadata, metadata format, RDF, resources description frame work, metadata standards

* 한국과학기술정보연구원 연구원

• 접수일 : 2003. 8. 21 • 최초심사일 : 2003. 8. 25 • 최종심사일 : 2003. 8. 28

I. 서 론

메타데이터란 상대적으로 최근에 출현한 용어이나 혼란스런 사용으로 그 의미하는 바가 사람마다 다르다. 이러한 상황에 대하여 린치 등(Lynch & et al., 1995)은 함축적으로 다음과 같이 기술하고 있다.

현시점에서 서술적 용어로서의 메타데이터는 너무 과도한 사용에 의해 평가절하하여(또한 사용분야나 문맥에 따라 너무나 많은 것을 의미하여) 복잡한 설명이 따르지 않으면 의미없는 것으로 우리에게 느껴진다. 바로 메타데이터라는 용어의 모호성 때문이다.

이러한 혼돈된 상태를 정리하여 데이(Day, 2001)는 간략하나 명료하게 메타데이터를 다음과 같이 정의하고 있다.

메타데이터란 문자적 의미로는 데이터에 관한 데이터 (data about data)라고 정의되나 일반적으로 다양한 업무를 지원하기 위하여 사용되는 정보자원에 대한 구조화된 데이터를 의미하는 것으로 해석된다. 메타데이터란 용어는 전자정보와 관련하여 처음 사용되기는 하였으나 그 후 정보자원에 대한 모든 표준화된 기술정보(standardized descriptive information)를 포괄하는 의미로 확장되었다 (emphasis added).

메타데이터란 용어는 데이의 정의처럼 초기에는 전자자료 특히 웹 자원 또는 네트워크 자원의 조직을 위한 포맷을 의미하였으나 점차적으로 그 영역을 확대하여 전자자료와 웹 자료는 물론 일반 정보자원의 기술(記述)과 조직을 위한 표준으로 자리매김해 가고 있다. 따라서 이 글에서는 데이의 정의와 같이 대상 매체형식에 상관없이 모든 정보자원에 대한 표준화된 기술(記述)정보를 포괄하는 포맷이라는 의미로 '메타데이터'를 사용한다. 특히, 이 글은 메타데이터 중 '문헌유사객체(document-like objects)¹⁾'를 대상으로 하는 메타데이터 형식에 대하여 논하고자 한다.

본 논문에서는 우선 개념적 혼란을 겪고 있는 메타데이터의 역할, 그 개념의 기원과 개념변화 및 혼란의 이유를 기초연구로서 간단히 살펴본다. 그리고 근래 들어 다른 메타데이터를 특정 포맷의 메타데이터로 변환하거나 여러 메타데이터 포맷을 하나의 포맷으로 통합하려는 움직임이 일고 있는데 대하여 (예. Chandler, Foley & Hafez, 2000; Duval et al., 2002; Library of Congress, 2001; Library of Congress, 2001a; McCallum, 2000; Society of American Archivists, 1999; St. Pierre & LaPlant, 1998) 그 원인과 목적이 무

1) Document-like object(DLO): 원래 문헌과 유사한 대상(객체)으로 정의되었다. 여기서 문헌유사객체이란 기본적으로 텍스트 기반 문서 또는 일반문서의 텍스트라는 속성 이외 다른 속성을 공유하는 객체를 의미한다. 예를 들면, 전자메일이나 스프레드 시트이다. 이러한 정의는 3차 더블린코어 워크숍(3rd DC workshop)에서 고정된 것, 즉, 모든 이용자에게 동일한 내용을 가지는 것이라고 특징지어진 모든 독립적 정보자원을 칭하는 것으로 확장되었다. 이들에 대한 예로는 텍스트, 이미지, 영화와 공연을 포함한 다. (source: <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml>)

엇인가 알아보고 그에 따른 문제점을 파악해 본다. 마지막으로 최근 이러한 움직임의 핵심요소로 대두되는 RDF (Resource Description Framework)와 메타데이터 레지스트리 (Metadata Registry)의 특성과 문제점, 그리고 이 둘의 관계를 살펴보고자 한다. 특히, RDF와 메타데이터 레지스트리는 상호보완적 관계인가 아니면 서로 상충되는 상이한 방법론인가를 밝히고 향후 발전방향을 제시하고자 한다. 이러한 과정은 메타데이터의 통합 검색이라는 측면을 염두에 두고 접근하였다.

II. 메타데이터: 역할, 용어의 출현 및 개념의 혼돈

전통적 의미에서나 인터넷 환경에서도 메타데이터의 중요한 역할은 정보검색을 원활히, 효과적으로 하기 위함이다. 정보검색에서는 검색시스템의 성능척도를 정확율과 재현율로 측정한다. 정확히 41년 전 클리블던(Cleverdon, 1962)은 유명한 클랜필드 실험 (Cranfield experiments)에서 정보검색시스템의 성능은 일반적으로 60%~90%의 재현율과 10~25%의 정확율에 머무른다고 결론지었다 (recited in Sparck Jones, 1981). 이러한 상황은 40여년이 지난 오늘날에도 큰 향상을 보이지 않고 있다. 결국 여전히 소수의 적합문헌을 얻기 위하여 많은 부적합 문헌을 살펴보아야만 한다.

반면, 인터넷 검색은 정확율에서 1%도 되지 않을 것으로 추정된다. 인터넷 검색엔진 www.google.co.kr에서 인명(人名) '심경'을 검색하면²⁾ 총 23,200 건의 검색결과가 나오고 그 결과 중 앞의 250개만을 보았을 때, 불교, 마음의 상태, 한의학, 농업에 관한 자료로 단순히 '심경'이라는 단어를 포함하는 모든 웹 문서가 검색되었다 (2002년 12월 29일 검색). 위에 언급한 정확율로 표현하자면 0%를 보이며 이는 검색엔진이 많은 부적합 문헌을 검색함을 나타낸다. 사실 검색엔진은 웹 문서의 주류를 이루고 있는 HTML 문서가 구조화되어 있지 않으므로 검색어를 저자, 서명, 주제 등 특정 필드에서만 검색할 방법은 없다. 만약 웹 문서에 대한 구조화된 정보가 있다면 정확율을 크게 향상시킬 수 있을 것이다. 그리고 재현율도 메타데이터에 의하여 향상될 수 있다. 예를 들면, 사이트가 HTML 이외의 이미지, 데이터베이스나 PDF(Portable Document Format)와 같은 문서를 가지고 있을 경우이다. 이외에도 검색엔진은 웹 문서의 모든 페이지를 색인하지 않는 것으로 알려져 있다. 검색엔진은 웹 문서의 깊이에서 대개는 상위 2~3계층만을 색인하므로 복잡하고 큰 사이트의 계층에서 하위에 속하는 중요한 문서(deep web 또는 invisible web이라고 불림)를 검색하지 못할 수 있다. 그러나 이러한 사이트도 해당 사이트에 메타데이터가 구축되어 있다면 검색엔진이 이들도 검색할 수 있을 것이다 (Cathro, 1997).

2) 본 예는 '저자명검색' 즉, known item search에 가까우며 앞선 '주제검색'과의 비교가 적절치 않으나 인터넷 검색의 비효과성(low-effectiveness)을 강조를 위하여 사용되었다.

기하급수적으로 증가하는 웹 문서에 대한 검색의 문제, 즉 효과적인 검색을 위하여 더블린 코어라는 13개 기본요소를 가진 메타데이터가 95년도 제1차 워크샵에서 제안되면서 (Weidel, S., Iannella, R. & Cathro, W., 1997) ‘메타데이터’라는 용어가 널리 사용되기 시작하였다. 그러나 실제 ‘메타데이터(metadata 또는 meta data)’라는 용어는 1969년 마이어스 (Jack E. Myers)에 의하여 데이터 셀을 효과적으로 기술하기 위한 개념으로 처음 만들어진 것으로 1973년에는 자신의 회사이름으로 사용되었다 (FOLDOC, 1997).

메타데이터라는 용어의 개념 혼란은 더블린 코어의 발표와 더불어 활발해진 메타데이터라는 용어의 사용, 그리고 더블린 코어의 발생을 전후하여 개발된 다양한 메타데이터 표준 또는 포맷이 대상으로 하는 객체와 깊은 관계가 있다. 이는 아래 <표 1>을 보면 명확해진다.

<표 1>메타데이터생성시기

메타데이터 명칭	최초 생성(제안)시기	대상 객체
TEI 헤더	1990년	전자문헌
Dublin Core	1995년	네트워크 상의 전자매체
EAD	1995-1998년	고문서
CDWA(Categories for the Description of Works of Arts)	1996년	예술작품
FGDC metadata	1998년	지리정보

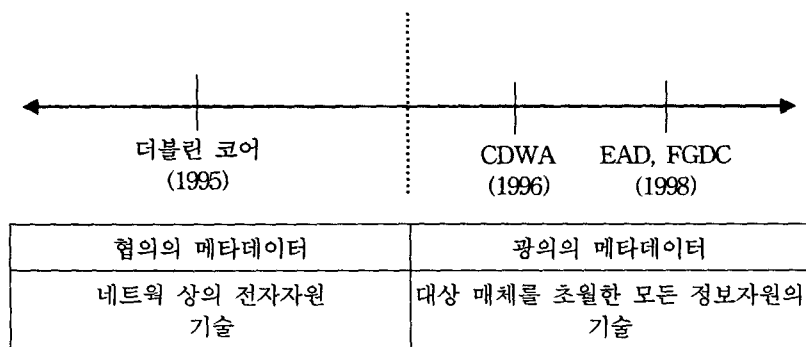
위 표에서 더블린 코어 이전에 출현한 메타데이터 포맷은 출판분야의 데이터 교환을 위하여 고안된 TEI(Text Encoding Initiative)뿐이다. 이를 제외한 다른 메타데이터들은 더블린 코어가 발표된 1995년을 시작으로 발표되었음을 알 수 있다. 이들의 발생이 더블린 코어에 의하여 유발되었는가 하는 인과관계는 명확치 않다. 그러나, 비록 이들이 모두 메타데이터라는 용어로 명명되었지만, 개별 포맷의 기술대상(記述對象) 객체는 1995년 제1차 워크샵 (OCLC/NCSA 메타데이터 워크샵)에서 발표된 더블린 코어처럼 기술대상자원으로 문헌유사객체로 간주되는 웹 자원(Web resources that could be thought of as Document-Like-Objects)의 목록을 위하여 개발된 것은 아니다.

애초 더블린 코어의 사용목적과는 달리 인터넷 자원 이외의 대상 기술을 목적으로한 ‘메타데이터’들이 비슷한 시기에 발표되고 이들 모두가 메타데이터라는 용어로 통합적으로 표현되면서 용어자체에 대한 혼란이 시작된 것으로 추정된다. 사실, 1995년 처음 발표 당시 네트워크 상의 전자자료를 대상으로 한다고 선언한 더블린 코어조차도 지금은 미디어 포맷에 상관없이 모든 종류의 자원(any kind of resource, regardless of the media format)에 관한 기초 기술정보 제공을 목적으로 한다고 명시하고 있어 (DCMI, 2003) 더

욱 혼란을 가중시키고 있다. 이러한 용어정의에 대한 변화와 혼돈은 UKOLN(UK Office for Library Networking) 메타데이터 그룹(1998)에서 검토 나열한 메타데이터 포맷만도 MARC를 포함하여 22개에 달하는 것을 보아도 충분히 짐작할 수 있다.

혼돈된 개념의 정리를 위하여 일부에서는 메타데이터를 범주화하여 기술적(技術的) 메타데이터 (technical metadata)와 비즈니스 메타데이터(business metadata)로 나누기도 한다 (예. Sherman, 1997; SAS, 2003). 전자는 데이터의 저장위치 및 접근방법 등에 관한 기술(記述)정보이며 후자는 데이터의 비즈니스적 문맥과 의미를 이해하기 위하여 작성되는 기술(記述)정보이다. 이 양분법은 경영정보라는 한정된 분야에서 정보시스템 관련 메타데이터와 의사결정 관련 메타데이터를 구분하기 위한 것으로 앞서 언급한 개념적 변이 또는 혼란을 정리하는 목적으로는 적절치 않다.

따라서 메타데이터라는 개념을 정리하면 '자원의 구조적 기술'이라는 공통요소를 가지나, 초기에는 기술대상 '자원'을 네트워크 상에 존재하는 '전자자료'에 한정하였으나 점차적으로 그 활용범위가 넓어지면서 '자원'을 '전자자료를 포함하여 매체를 초월한 객체'로 확장 적용하고 있음을 알 수 있다. 이러한 개념적 변이는 <그림 1>과 같다.



<그림 1> 메타데이터 개념의 변이

결론적으로 메타데이터란 대상 매체를 초월한 모든 정보자원의 구조적 기술 포맷 또는 기술내용이라고 정의할 수 있다. 그러나 이에 대한 해석은 앞서 인용한 린치 등(Lynch & et al., 1995)이 지적한 것처럼 '정보자원'에 대하여 사용분야나 문맥에 따라 조심스런 접근이 필요할 것이다. 예를 들면, 문헌정보 분야에서는 DCMI에서 정의한 문헌유사객체를 정보자원으로 보는 견해가 일반적이다. 역설적으로 말하면 사용하는 경우에 따라 문헌유사객체가 아닌 객체에 대한 구조적 기술 또한 메타데이터라고 할 수 있다.

III. 새로운 메타데이터 포맷의 발생

오랫동안 문헌정보학 분야에서 자원기술을 위한 유일한 표준으로 사용되던 MARC 포맷은 인터넷 자원의 폭증으로 이를 수용하기 위한 수정이 불가피하게 되었고 또한 새로운 형태와 접근방식을 가진 웹 자료의 기술에 적합한 포맷이 생겨나기 시작하였다. 이와 같은 상황을 김태수, 김이겸(1998)은 다음과 같이 요약하였다:

웹 자원을 저록으로 표현하기 위한 방안은 크게 두 가지 흐름이 있다. 하나는 기존의 MARC형식에서 네트워크 자원의 소재와 접근에 관한 정보를 특정 필드(856)로 수용하여 모든 유형의 자료를 통합하는 방안이다 다른 하나는 분산구조라는 관점에서 메타데이터의 구조를 특정 형식으로 제한하지 않고, 다양성을 인정하는 관점이다. 기본적으로 하나의 메타데이터 형식이 독점적으로 사용될 수 없고, 또한 특정 형식이 무한정 확장될 수도 없다는 관점에 배경을 두고 있다 (p.3).

웹 자원을 표현하는데 있어 기술형식의 다양성을 인정하는 관점의 대표되는 메타데이터 포맷은 더블린 코어이다. 그러나, 실제 확장된 더블린 코어의 15개 요소를 살펴보면 MARC 포맷이 지원하지 못하는 요소는 없다 (아래 <표 2>참조).

<표 2> 더블린 코어 요소와 대응되는 MARC 태그

더블린 코어 요소	MARC 포맷 태그
1. Title	130 Uniform title; 210 Abbreviated title; 211 Acronym or short title; 212 Variant access title; 214 Augmented title; 222 Key title; 240 Uniform title; 241 Romanized title; 242 Translation of title by cataloging agency; 243 Collective uniform title; 245 Title statement; 246 Varying form of title; 247 Former title or title variations 외 4XX Series statement, 7XX Added entries
2. Creator	100 Personal name; 110 Corporate name; 111 Meeting name; 245 \$c Statement of responsibility, etc.
3. Subject	600 Person name (subject); 610 Corporate name (subject); 611 Meeting name (subject); 630 Uniform title (subject); 650 Topical term (subject); 기타 65X
4. Description	520 Summary, etc.; 505 Table of Contents
5. Publisher	260 \$c Name of publisher
6. Contributor	245 \$c Statement of responsibility, etc.; 700 Personal name; 710 Corporate name 등 7XX의 Added entries
7. Type	008 23 th Form of item; 008 25-27 th Nature of contents; 650 \$v Form subdivision; 655 Genre/Form
8. Date	이하 생략

위 표는 더블린 코어와 MARC 포맷 간의 변환테이블³⁾ 제공이 목적이 아니라 두 개의 메타데이터 포맷 간에 상호변환이 가능하며, 따라서 어떤 객체를 기술하기 위하여 반드시 이 중 하나의 메타데이터 포맷을 필요로 하지는 않는다는 것을 시사한다.

그렇다면, 왜 새로운 메타데이터 포맷이 발생하였을까? 특히, 소위 문헌유사객체(document-like objects)를 기술함에 있어 가장 널리 사용되고 있고 성숙된 포맷을 가지며 (Dempsy & Heery, 1998) 질적으로도 우수한 MARC포맷 이외의 다른 메타데이터 포맷이 필요했던 것일까?

이 질문에 대한 답변은 두 가지로 할 수 있다. 앞선 메타데이터 정의에 의하면 새로운 메타데이터 포맷의 발생은 크게 두 가지 이유로 나뉜다: 첫째, 기하급수적으로 증가하는 웹 자원에 대한 기술과, 둘째, 웹 자원 또는 전자자원이 아닌 전통적 매체에 대한 기술이다.

1. 웹 자원을 위한 메타데이터 포맷의 발생

앞서 살펴본 바와 같이 초기 웹 상의 전자자원을 기술하기 위한 포맷으로서 더블린 코어의 기술요소는 기존 MARC 포맷으로 충분히 지원이 가능하다. 사실, 더블린 코어의 발생은 MARC 포맷의 한계성 문제가 아닌 레코드 생성의 경제성과 관련이 있다.

상술하면 더블린 코어의 출현은 인터넷의 보급과 HTML 기반 WWW(World Wide Web) 문서의 급증과 이들에 대한 색인부재와 결과적으로 적절한 검색방법의 결여가 직접적인 원인이다. 웹 자원은 나름대로의 색인을 가지고 있으나 웹 문서의 색인은 구조화되어 있지 않아 대상 주제범위를 확대해 감에 따라 검색결과와 주제영역 간에 어의표류(semantic drift) 문제가 구조화된 색인보다 심각해진다. 이러한 문제극복을 위하여 웹 자원의 기술을 위한 메타데이터 포맷이 요구되었다.

이 상황에 대처하는 데에 MARC가 적합하지 않았던 이유는 포맷 자체의 결함이나 한계성보다는 MARC 레코드 생성 단가(單價)였다. 일반적으로 MARC 레코드는 생성을 위하여 전문인력 투입이 필요하며 생성 및 유지관리에 많은 시간이 소요되므로 단지 중요한 자료에만 적용이 되어왔다. 즉, 고비용의 MARC 레코드를 양적(量的)으로 방대한 반면 질적(質的) 제어가 이루어지지 않은 웹 자원의 기술에 적용하기에는 비용 대비 효과면에서 적절치 않았다. 따라서 웹 자원의 검색효과 향상을 위한 구조화된 메타데이터의 필요성과 기존 MARC 포맷의 비경제성이라는 상충되는 현상을 완화하는 차선택이 필요하였던 것이다. 이렇게 발생한 더블린 코어 포맷은 웹 자원 고유의 색인보다는 정보가 많으나 정식 목록 레코드, 즉, MARC 레코드보다는 덜 완벽한 레코드 생성을 위한 포맷으로서 축소된 노력으로 더 많은 대상의 기술확보를 위한 기반이라고 볼 수 있다.

3) 더블린 코어와 MARC포맷 간의 변환테이블은 Dublin Core/MARC/GILS crosswalk (<http://lcweb.loc.gov/marc/dccross.html>)을 참조하라. 위 표는 MARC의 고정장 필드까지를 포함시킨 점이 Crosswalk와는 구분된다.

그러나 웹 자원 기술을 위한 MARC 포맷의 부적합성을 해결하기 위하여 제안된 더블린 코어도 자체적 문제점을 가지고 있다. 캐드로(Cathro, 1997)는 더블린 코어 표준의 제정 이유와 문제점에 관하여 다음과 같이 간략히 설명하고 있다:

더블린 코어는 1995년과 1996년 사이에 정보자원의 검색효과 향상의 필요성에 부응하여 만들어졌다. 여기서 정보자원이란 특히 웹을 의미한다. 더블린 코어는 도서관, 고문서관, 정부와 온라인 정보 발행자들의 사용을 취지로 개발된 포괄적인 메타데이터 표준이다... 더블린 코어 사용자들에 의하여 어의(15개 필드의 내용에 대한 규칙)와 구문(필드의 구조화와 표현에 대한 규칙)에 대한 여러 가지 우려가 표시되었다. 두 가지 측면에서 모두 다 더블린 코어는 불안정하며 합리적인 빠른 변화가 일어날 것이다.

위 설명에서 보듯 더블린 코어는 초기 웹 문서를 생성하는 저작자에 의하여 자원에 관한 기술(記述)이 작성되도록 고안된 것이다. 따라서 비전문가의 사용이 전제되어 포맷의 간략성과 사용성(usability)은 필수조건이었다는 점을 감안하면 캐드로의 지적은 당연한 결과일 수도 있다. 어쨌든, 비전문가에 의한 메타데이터 생성은 기하급수적으로 증가하는 웹 자원 전체의 조직에 현재 도서관 목록업무를 적용하는 것은 불가능하기 때문에 (McCallum, 2000) 그 발생이 필수 불가결하였다고 볼 수 있다. 더블린 코어의 발생은 그렇다고 하더라도 다른 메타데이터 포맷의 상황적응 논리는 무엇일까?

2. 일반 자원을 위한 메타데이터 포맷의 발생

더블린 코어는 발생 당시 기술대상이 웹 자원에 한정되었으나 기존 MARC 포맷의 활용이 가능하다고 생각되었던 문헌유사객체에 대해서도 MARC 이외에 다양성을 추구한 새로운 포맷들이 속속 개발되었다.

새로운 메타데이터 포맷 구성의 이유는 더블린 코어와는 달리 비용 대비 효과 측면보다는 대상객체를 MARC로 기술하는데 따른 물리적, 논리적 한계점 때문이다. 예를 들면, EAD (Encoded Archival Description)는 초기 더블린 코어와는 달리 전자자원을 대상으로 하지 않고 전통적 매체에 기록된 고문서 자료의 기술을 위한 포맷이다. EAD 개발자들은 새 포맷의 개발 이유로 MARC 포맷의 최대허용 레코드 사이즈, MARC 포맷의 평면적 특성이 자신들의 계층적 구조정보에 적용하기 어려움, 그리고 MARC 채용 시 필요 장비 운영과 프로그램 개발에 필요한 대단위 예산소요라는 비용문제 등 세 가지를 들고 있다 (Dooley, 1998). 이와 같이 MARC 포맷의 물리적 한계, 즉 포맷의 구조적, 기술적(技術的) 단점을 극복하고자 생성된 메타데이터 포맷은 EAD 이외에도 FGDC 메타데이터 등 다수이다.

여기서 MARC 포맷의 물리적 논리적 한계가 무엇인가에 대하여 부연할 필요가 있다. MARC 포맷의 물리적 한계 중 구조적 문제로 자주 언급되는 것은 레코드 최대길이, 필

드 최대길이, 한 레코드 내의 최대 필드 수 등이며 기술적 문제로 제기되는 것은 MARC 포맷의 계층구조를 지원하기 어려운 평면성이다 (예. Chandler, Foley & Hafez, 2000; Dooley, 1998). 결국 서지기술을 위하여 충분한 규모라고 제안되었던 MARC의 규격은 보다 복잡해지고 다양해진 최근에 요구되는 자원기술의 필요성을 충족시키기에는 단위 용량자체가 부족하여 물리적 한계를 보이고 있는 것이다. 또한, 기술적 문제란 MARC의 근간을 이루는 데이터베이스라는 기반 기술을 의미한다. 이 기술이 계층구조의 정보표현에 적절치 않다는 것이다. 정보의 계층구조란 기존 도서관 개념을 도입하여 설명하면, 집서 수준(collection level), 하위구조(subunit), 그리고 개별자료(item)의 구성요소의 관계를 의미한다. MARC 포맷을 사용하여 이와 같은 관계를 조직하자면 다수의 레코드를 생성하고 이들간의 연계 등을 통하여 계층구조를 형성하여야 하는데, 데이터베이스 기반의 MARC 포맷은 이러한 계층구조를 지원하기에 제한적이라는 것이다 (Dooley, 1998; Miller, 2000). 이러한 지적은 근래 XML와 SGML이라는 틀의 출현이 적지 않은 역할을 했다. 이들은 하나의 레코드 내에서 계층구조를 표현할 수 있도록 하고 있다.

이와 같이 MARC 포맷의 물리적 제한점으로 인한 새로운 포맷의 생성과 더불어 근래에는 MARC 포맷에 대한 논리적 구조의 문제점도 활발히 지적되고 있다 (Miller, 2000; Johnson, 2001). MARC 포맷 구조의 비논리성이란 예를 들면, 발행일의 동일 레코드 내에서 반복적 표현(008 07-10, 11-14와 태크 260 \$c) 등이며 이러한 반복적 표현이 오히려 혼란과 오류의 원인이라는 것이다.

그러나 개별 상황의 다양성을 인정하여 생성된 메타데이터 포맷들은 독자적인 필요를 만족시킬 수는 있었으나 데이터 공유 또는 교환이라는 측면에서는 상황을 MARC 포맷이 생성되기 이전의 상태로 돌려놓는 결과를 초래했다.

IV. 특정 메타데이터 포맷의 기존/다른 형식으로의 변환

메타데이터 표준의 종류가 증가함에 따라 각 표준에 맞추어 메타데이터를 제공하는 것은 점차 반복적이며 시간을 요하고 어렵게 되었다. 이러한 메타데이터를 생성하고 유지하는데 필요한 시간을 최소화하고 폭넓은 이용자 층에게 유용성을 최대화하기 위하여 관련된 분야의 메타데이터 표준들을 하나의 표준을 통하여 접근할 수 있는 메타데이터가 생성되고 유지될 필요성이 생겼다 (St. Pierre & LaPlant, 1998). 결국 이러한 필요성은 다양성의 통일이라는 오래 전에 해결되었다고 믿어졌던 상황과의 재조우(再遭遇)를 의미하는 것이다. 따라서 유사한 분야의 복수 메타데이터를 하나의 메타데이터 표준으로 접근시키고자 하는 노력이 다각적으로 시도되고 있다. 이러한 노력들은 대략 두 가지 방향으

로 요약된다.

첫째, 특정 메타데이터 포맷의 기존/다른 형식으로의 변환과

둘째, 여러 메타데이터를 포괄할 수 있는 범용적 구조(framework)를 통한 통합이다.

1. 특정 메타데이터 포맷의 기존/다른 형식으로의 변환

새로운 메타데이터의 생성으로 발생한 문제점은 이용자의 새로운 포맷에 대한 혼란, 관리의 어려움 및 복수 시스템의 불가피한 사용을 들 수 있다. 이에 대한 예로 FGDC (Federal Geographic Data Committee) 메타데이터를 들 수 있다. 1998년 FGDC는 클리어링 하우스의 National Biological Information Infrastructure (NBII) 부문에 제출된 50% 이상의 검색문이 검색결과 0건을 기록하였다. 1년 간에 걸쳐 문제 추적분석 후 FGDC가 취한 두 가지 해결방안 모두는 메타데이터 표준에 관한 것으로 클리어링 하우스 이외의 다른 시스템에도 호환되는 보다 폭넓게 사용되는 메타데이터 표준 (즉, MARC와 더블린 코어)으로 현재 작성된 그리고 향후 구축될 메타데이터의 변환 및 보다 관리가 용이한 메타데이터 포맷 지원을 위한 시스템 모델의 변경이었다 (Chandler, Foley & Hafez, 2000). 또한 OCLC의 지원으로 인터넷 자원의 메타데이터를 도서관들이 생성하고 공유하자는 취지에서 시작된 CORC (Cooperative Online Research Catalog)는 더블린 코어와 MARC21을 통합하여 전자자원과 전통매체의 메타데이터를 하나의 시스템에서 제공하기 위한 노력이다 (Hickey, 1998). CORC는 질 높은 MARC포맷으로의 변환이 이미 널리 보급 사용되는 도서관자동화 시스템에 다른 메타데이터의 통합을 가능케 하기 때문에 특히 가치가 있다 (Demsey & Heery, 1998).

위의 예에서 시사하듯 고유한 필요성에 의하여 확립된 각종 메타데이터 포맷이 결과적으로 메타데이터 생성의 목적인 원활한 검색과 관리의 두 가지 측면 모두에서 문제를 초래하고 있음을 알 수 있다. 사실, 메타데이터 표준이 해당 주제 또는 분야의 필요성에 의하여 다양해지는 것은 바람직하나 폭 넓은 이용자 집단이 손쉽게 이들에 접근하고 이용하기 위하여는 메타데이터 포맷 간의 공통된 요소가 존재하여야 하며 이는 많을수록 바람직하다. 이러한 전제에서, 만약 자동화된 메타데이터 포맷 요소 간의 상관표 (matching tables 또는 crosswalks)가 존재한다면 검색엔진들이 최소한 특정 부류(given family)의 메타데이터들의 콘텐츠 검색에 작동 (function)할 것이다 (St. Pierre & LaPlant, 1998).

그러나 메타데이터 표준은 일련의 데이터 요소와 그 의미를 정의한 것으로 개별 요소는 명칭에 사용된 용어가 서로 상이하며 그 의미의 고유성을 가지고 있다. 이러한 의미의 문제와 공통 요소의 상관관계를 인간노력으로 정의하여 하나의 공통된 형식으로 통일하는 것은 가능하며 이미 일어나고 있다. 여기서 공통 포맷으로 가장 널리 사용되는 것은 오랜 역사와 최고의 레코드 수를 이미 보유하고 있는 MARC이다. 그럼에도 불구하고 특

정 메타데이터를 기존에 개발되어 널리 사용되는 포맷으로의 변환에 따르는 문제는 작업 절차의 어려움보다는 결과물의 만족스럽지 못함이다. 이들의 변환은 상호 요소 간의 상관표를 구성하여 대응되는 요소로 옮기는 비교적 간단한 절차가 사용된다. 하지만 일견 간단해 보이는 절차의 배경에는 각 요소의 기술법칙 및 질적 문제와 상호 대응되는 요소의 결핍 또는 그 반대현상으로 불가피한 정보손실이라는 구조적 문제의 두 가지 바람직하지 않은 측면이 있다.

하나 이상의 메타데이터 포맷을 변환 또는 통합하였을 때 발생하는 질적 문제는 웹자원 기술을 위하여 개발한 메타데이터 포맷의 콘텐츠를 MARC와 같은 정식 메타데이터 포맷으로 변환할 경우이다. 예를 들면, 더블린 코어는 MARC 형식보다 기술수준(description level)이 낮아 MARC 형식의 저자필드는 개인저자와 단체로 구분되는데 반해서 더블린 코어에서는 이러한 저작구분을 적용하지 않기 때문이다. 뿐만 아니라 데이터 요소의 구축에 필요한 규칙이나 적용되는 프로토콜도 아직 개발되지 않은 상황이다 (김태수, 김이경, 1997. p. 23). 또한 MARC 포맷 레코드를 더블린 코어 포맷으로 변환하면 전자(前者)의 여러 필드가 후자(後者)의 단일 요소에 합쳐져 이를 역방향으로 기계적 재변환은 불가능한 구조적 문제도 발생한다.

특정 메타데이터 포맷의 다른 포맷으로의 변환은 해당 메타데이터가 가지고 있는 콘텐츠 요소의 종류와 양에 영향을 받는 구조적 문제의 다른 예로 FGDC 메타데이터 (정식명칭은 Content Standard for Geospatial Metadata: FGDC-STD-001-1998)를 MARC21과 더블린 코어로 변환하는 프로젝트를 들 수 있다 (Chandler, Foley & Lafayette, 2000). 정보기술(情報記述)의 양적 측면에서 MARC21 포맷은 99,999 바이트가 이론적 최대 레코드 크기이다. 그러나 이를 기준으로 구현된 개별 도서관 자동화시스템들은 최대허용 레코드 크기가 서로 다소 다르다. 예를 들면, 국제적으로 가장 많은 MARC 레코드를 보유한 OCLC의 WorldCat이 허용하는 최대 레코드 크기는 4,096 바이트이며 가변장 필드의 최대 크기는 1,230 바이트이다. 또한 최고 50개까지의 가변장 필드를 허용한다. 반면 National Biological Information Infrastructure (NBII)의 FGDC 메타데이터를 조사한 결과는 다음 표와 같다.

〈표 3〉 FGDC 메타데이터 특성 통계

	레코드길이	레코드내최대필드길이
평균	6,792바이트	2,125바이트
메디안	6,474바이트	1,258바이트
대상466레코드중최대레코드	28,042바이트	9,525바이트
4096바이트보다큰레코드수	343레코드	
1230보다긴필드를가진레코드 수		236레코드

위 표에 나타난 수치들은 명확히 FGDC 메타데이터를 WorldCat 시스템으로 변환하는 데는 무리가 있음을 보여준다. 이 조사결과에서 74%의 레코드가 최대 레코드 길이를 초과하며 51%의 레코드가 최소 하나 이상의 최대 필드 길이를 초과하는 필드를 가지고 있는 것으로 밝혀졌다. 또한 정보의 종류에서 FGDC 메타데이터는 300 종이 넘는 엘리먼트를 가지고 있어 이들을 모두 MARC 포맷으로 변환하는 것은 무리로 밝혀졌으며 특정 기준에 따른 선택적인 변환이 불가피하였다.

기존 MARC 포맷의 허용 레코드 또는 필드 크기에 대한 제약 및 추가적인 한계점은 새로운 메타데이터 포맷 고안의 원인이었고 이들의 다양성은 다시 통합 또는 통일이라는 문제를 야기하였다. 이는 다시한번 MARC 포맷으로의 회귀를 추구하게 되었으나 이 또한 위에 설명한 것처럼 만족스럽지 못하다.

따라서 개별 메타데이터의 특징을 보존하기 어려운 기존 메타데이터 포맷을 이용한 변환이나 통합과는 별도로 제3의 범용적 구조를 사용한 기술표현 방법이 제안되고 있다.

2. 범용적 구조를 통한 통합

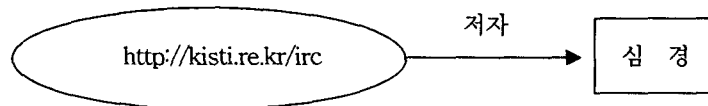
범용적 구조를 이용한 통합이라고 하는 것은 개별 메타데이터 포맷의 특징을 인정하면서 다양성으로 인한 데이터의 분산과 중복노력을 극복하고자 하는 접근방식을 의미한다. 이중 하나는 서로 다른 메타데이터 포맷을 특정 포맷으로 변환하여 통합함으로써 발생하는 문제점을 해결하기 위하여 기존 포맷을 인정하면서 이들을 하나로 묶을 수 있는 틀을 제공하는 것이다. 이는 RDF(Resource Description Framework)라는 메타데이터 통합에 적합한 틀의 사용이다. 다른 하나는 서로 다른 메타데이터가 다양성을 가질 수 있으나 사전 또는 사후에 이러한 포맷에 대한 '등록'을 실시하여 (1) 새로운 포맷의 생성시 공통된 요소에 대하여 최대한 과거 포맷에 접근하도록 하고, (2) 이미 존재하는 포맷에 대하여도 서로 공통점과 차이점을 등록정보를 참고할 수 있도록 하는 것이다. 이를 메타데이터 레지스트리(Metadata Registry)라고 한다.

V. RDF를 이용한 메타데이터 통합

기존 방식과는 별도로 다양한 메타데이터 포맷들의 고유성을 인정하면서 통합할 수 있는 기술언어(記述言語) 또는 기술의 틀로서 RDF(Resource Description Framework)가 검토되고 있다 (예. 김태수, 김이경, 1998; Kokkelink & Schwanzl, 2002).

RDF는 웹 상의 자원에 관한 정보를 표현하기 위한 언어이나 웹자원이라는 개념을 일반화함으로써 웹에서 직접 검색이 되지 않을지라도 표현할 수 있는 자원기술을 위한 틀이다 (W3C RDF Core Working Group, 2003). RDF는 자원기술 틀로 다양한 메타데이터 형식과의 연계를 통한 통합기술(統合記述) 틀을 제공하여 스키마 참조를 기초로 상이한 형식의 메타데이터를 RDF로 수용할 수 있다.

RDF⁴⁾는 RDF 선언문(statement)이며 주제(subject), 술어(predicate)와 목적(object)의 기본구조로 구성된다. 이들은 각각 자원, 속성, 그리고 속성값으로 해석된다. 이러한 구조를 활용하여 간단한 형태의 심경은 자원인 RDF입문서(<http://kisti.re.kr/irc>)의 저자이다라는 문장을 표현하면 다음과 같은 RDF 그래프로 표현할 수 있다:



〈그림 2〉 간략한 RDF 선언문의 그래프

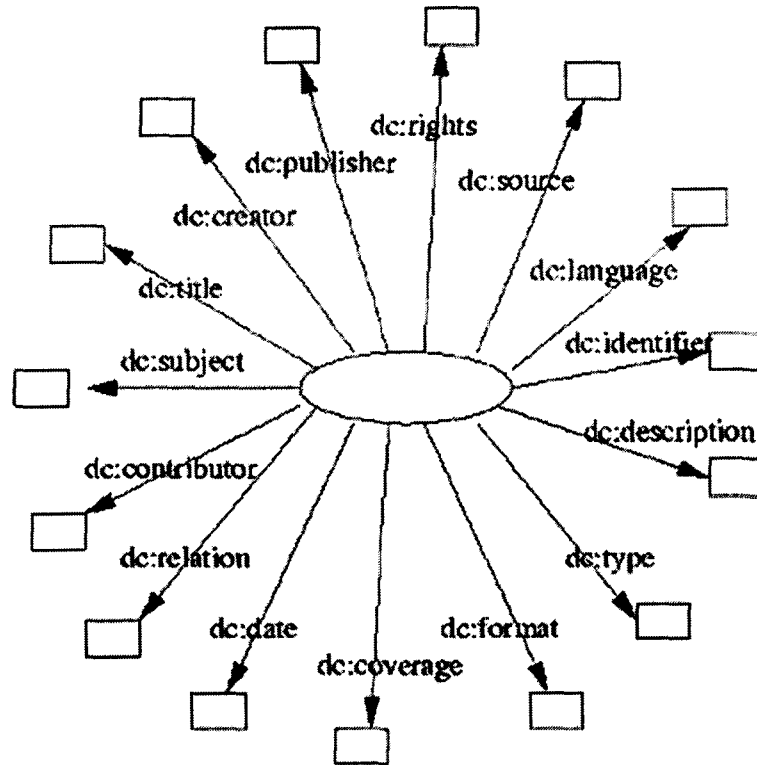
위 RDF 그래프는 다음과 같이 해석이 가능하다.

〈표 4〉 〈그림 2〉 그래프의 설명

주제(자원)	http://kisti.re.kr/irc 라는URL로표시된 “RDF입문서”
술어(속성)	저자
목적(속성값)	심경

위 예는 가장 간단한 형태의 RDF 모델을 표현하고 있으나 RDF 입문서라는 자원이 웹 상에 존재하며 URL을 가지고 있으므로 이를 참조하여 그 책의 ‘저자’라는 속성(즉, 관계)을 그 속성값인 저자명 ‘심경’과 연계하여 표현이 가능하다는 개념을 보여주고 있다. 위 그래프를 조금 응용하여 확장하면 다음과 같은 더블린 코어의 요소를 나타내는 그래프를 만들 수 있다.

4) 김태수(1999). 네트워크자원과 메타데이터. 정보관리연구, 30, pp. 1-26.은 RDF 이해를 위하여 간단명료한 설명을 제공한다. RDF에 대한 기본개념, 구문, 어휘기술방법 및 RDF를 이용한 응용프로그램의 개요는 RDF Primer (W3C RDF Core Working Group, 2003)를 참조.



〈그림 3〉 더블린 코어 기술요소의 RDF 그래프

위 그래프가 의미하는 것은 타원형 부분이 비워져 있으나 이를 우리가 원하는 대상자 원을, 예를 들면, 'RDF 입문서'라고 간주하였을 때 그 메타데이터를 더블린 코어의 기술 요소를 적용하여 해당 데이터를 박스 안에 넣을 수 있다고 해석하면 된다. 이에 대한 좀 더 상세한 설명을 위하여 위 그래프를 다음에 표현된 RDF/XML이라는 구문을 보면 더 명확해 진다.

1. <?xml version="1.0"?>
2. <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 - 3. xmlns:dc="http://purl.org/dc/elements/1.1/">
4. <rdf:Description>
5. <dc:creator>a</dc:creator>
6. <dc:contributor>b</dc:contributor>
7. <dc:publisher>c</dc:publisher>
8. <dc:subject>d</dc:subject>
9. <dc:description>e</dc:description>

10. <dc:identifier>f</dc:identifier>
11. <dc:relation>g</dc:relation>
12. <dc:source>h</dc:source>
13. <dc:rights>i</dc:rights>
14. <dc:format>j</dc:format>
15. <dc:type>k</dc:type>
16. <dc:title>l</dc:title>
17. <dc:date>m</dc:date>
18. <dc:coverage>n</dc:coverage>
19. <dc:language>o</dc:language>
20. </rdf:Description>
21. </rdf:RDF>

(RDF/XML 구문의 각 행 앞에 일련번호는 설명을 위하여 붙인 것임)

위 그래프에 표시된 'dc:contributor'라는 속성의 의미는 그래프에 상응하는 RDF/XML 구문의 3행과 6행에 의하여 의미를 명확히 할 수 있다. 3행을 네임스페이스(name space) 선언부라고 하여 이는 선언부 이후에 출현하는 구문에 'dc:'라는 접두사는 3행에 정의한 URI(Universal Resource Identifier)에서 가져온 것이라는 것을 의미한다. 따라서 6행의 'dc:contributor'라는 태그는 'http://purl.org/dc/elements/1.1/'이라는 웹 상의 위치에 정의된 더블린 코어라는 메타데이터 포맷의 'contributor' 속성요소를 나타내는 축약형 표시이다. 간략히 설명하면 이와 같은 방법으로 RDF는 다른 곳에서 정의된 메타데이터 포맷을 재생성하지 않고 단순히 '참조'함으로써 표현할 수 있는 기술(記述)모델을 제공한다. 이와 동일한 방법으로 우리는 MARC나 다른 메타데이터 포맷을 RDF를 이용하여 기술할 수 있다.

이와 같이 상이한 형식의 메타데이터를 그대로 수용한다는 점에서 위에서 언급한 상호 요소 간의 상관표(matching tables 또는 crosswalks)가 가지는 단점을 극복할 수 있으나 반면 RDF는 기술 요소 간의 어의적 연계성의 문제가 아직은 미해결로 남아있다 (김이겸, 1999). 예를 들어, 더블린 코어에서 'reator'로 정의한 요소가 MARC에서는 'author'란 다른 용어를 사용하나 그들이 의미하는 바는 기본적으로 동일한 대상이다. 그러나 RDF에서는 이러한 어의적 표현을 원 의도대로 표현하거나 이해할 수 있는 방법이나 정의를 마련하고 있지 않다. 즉, RDF 스키마명세는 서명이나 저자와 같은 기술요소를 직접 제공하는 것이 아니라 이들을 정의하는 수단을 제공할 뿐이다.

이러한 RDF의 개념과 개념 간의 관계 정의를 초월하여 그러한 의미구조를 담을 수 있는 방법으로 이러한 개념관계에 의미구조를 부여하기 위한 온톨로지(ontology)의 적용(김

이란, 2001)과 RDF를 이용한 개별 메타데이터 스키마를 포괄하는 상위 스키마를 활용한 방법(김이겸, 1999)이 연구되었다.

먼저 김이관(2001)은 위의 문제를 다양한 메타데이터 포맷을 통합하여 실제 구현된 시스템의 검색을 위한 방법으로서 보다는 RDF의 어의적 문제점에 대한 개념적 해결방안을 모색하였다고 할 수 있다. 우선, RDFS(RDF Schema)가 의미구조를 반영하지 못하므로 발생하는 문제점으로 다음 세 가지를 지적하였다:

첫째, 메타데이터나 데이터베이스 스키마를 구축하는 사람마다 개념을 개별적으로 정의하여 그 개념 간의 관계와 구조가 달라지며,

둘째, 스키마는 고유한 URI로 식별되어 네임스페이스를 통해 정확히 하나의 대상자원이 참조될 수 있도록 할 수 있으나 하나의 스키마를 다른 이름으로 지칭하여 참조한 두 스키마는 같은 대상자원을 사용함에도 불구하고 그들이 동일한 대상임을 식별할 수 없고,

셋째, 스키마의 새로운 버전에는 고유한 URI가 주어지므로 각기 다른 스키마로 간주된다.

이러한 문제점을 해결하기 위한 방법으로 제한된 영역 온톨로지를 개발하여 적용하였다. 온톨로지는 분야에 따라 여러 가지 의미와 종류가 있으나 이 논문에서 사용한 것은 그 기본요소를 영역의 주요 개념과 그 개념 간의 관계를 정의한 것이다. 이 연구는 앞서 언급한 바와 같이 RDF의 어의적 제한점을 극복하고자 온톨로지 기법 적용을 제안하였으나 메타데이터 포맷의 통합이라는 세부적인 주제에 초점을 맞추고 있지는 않으며 실제 검색을 위한 연구는 진행하지 않았다.

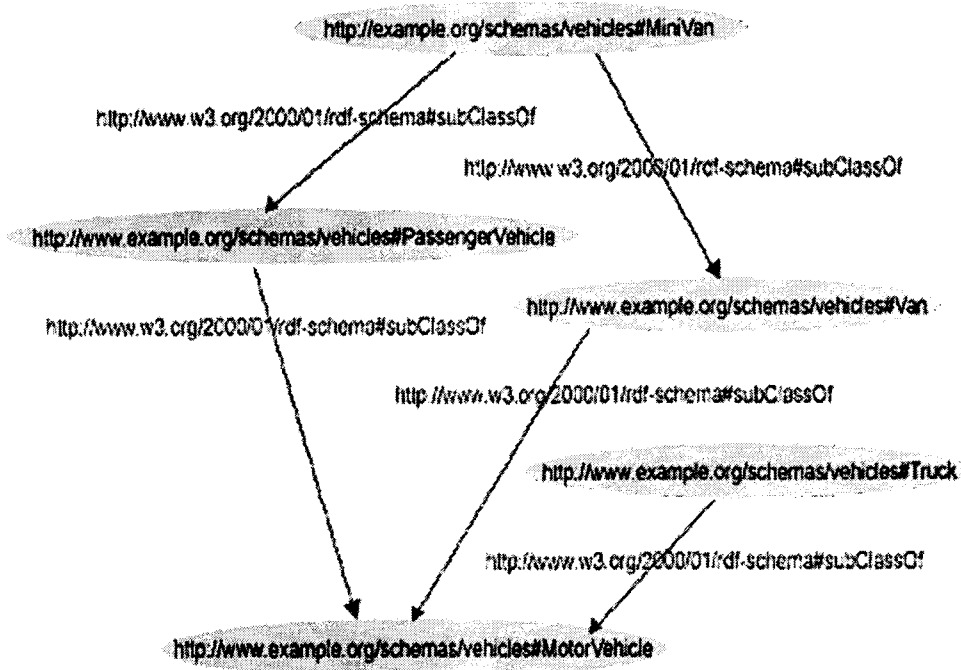
반면, 김이겸(1999)은 RDF를 활용한 메타데이터 통합의 문제를 직접 해결하고자 하였다. 특히, 과거 관련 연구에서 다루어지지 않았던 이종 메타데이터 포맷의 통합검색에 초점을 맞추었다. 이 논문은 KORMARC와 더블린 코어를 대상으로 메타데이터를 통합하기 위하여 3개의 스키마를 구성하였다.

먼저 두 메타데이터 형식 간의 공통요소를 추출하여 상이한 형식의 기술요소를 집단화하여 공통요소에 속한 기술요소 집단은 어의적인 연계성을 가지며, 또한 RDF 구문에서는 사용되는 기술요소는 메타데이터 스키마에 정의되기 때문에 어의적인 연계성 확보 역시 스키마를 통해 달성된다고 가정하였다. 그리고 추출한 두 메타데이터 공통요소는 모든 형식의 스키마 설계에 공동으로 활용될 수 있도록 별도의 스키마 설계에 반영을 하였으나 이들은 메타데이터 기술에 직접적으로 활용되는 기술요소가 아니므로 속성이 아닌 클래스로 정의하였다.

즉, 개별 메타데이터에 대한 스키마는 메타데이터의 기술요소를 사용하여 MARC와 더블린 코어 포맷의 것을 각각 설계하고 이들을 포괄하는 공통 요소를 앞서 언급한 상위 스키마에 포함시킨 것이다. 이 방법에서 중요한 사항은 클래스를 사용하여 정의했다는 것이다.

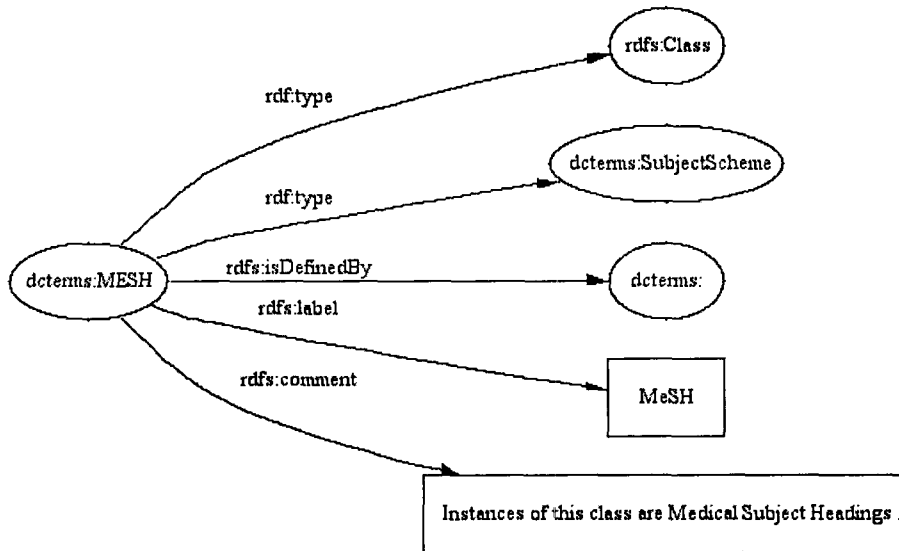
RDF에서 클래스란 유사한 타입이나 부류(즉, 카테고리)를 포괄하는 의미이다. 예를 들면, 여러 종류의 자동차에 관한 정보를 제공하기 위하여 RDF를 사용한다고 가정하자.

RDF 스키마에서는 먼저 자동차라는 대상의 부류를 표현하는 클래스가 필요하다. 그리고 특정 클래스에 속한 자원을 그 클래스의 인스턴스라고 하는데 인스턴스로는 승용차, 트럭 또는 개별 모델인 캐딜락, 멀세이디즈 등이 속할 수 있다. 여기서 가지는 의도는 이 클래스에 속하는 인스턴스가 자동차라는 자원이라는 것이다. 이를 좀 더 상세히 전개하고 싶을 경우, 예를 들어, 특수 차량이라는 개념을 기술하고자하면 RDF의 하위 클래스(subclass)라는 개념을 사용한다. 이를 그래프로 예시하면 다음과 같다:



〈그림 4〉 RDF의 클래스, 하위클래스 구조

이러한 특성을 가진 클래스를 이용하여 공통요소의 스키마를 메타스키마(meta_schema)로 정의하여 사용하였다. 따라서 메타스키마의 클래스에 속한 인스턴스를 예시하면 아래 그래프와 유사한 개념을 사용한 것이다.



〈그림 5〉 RDF의 클래스와 인스턴스 관계 예시

앞 그래프는 아래 표의 더블린 코어 용어 RDF 스키마를 RDF 스키마 클래스로 표현한 것이다.

〈표 5〉 〈그림 5〉에 나타난 더블린 코어 요소

더블린코어요소	요소인코딩스킴
Subject	LCSH
	MeSH
	DDC
	LCC
	UDC

이와 같은 정의를 이용하여 메타스키마에서 예를 들면 ‘TitleClass’를 정의하고 그에 해당하는 KORMARC의 ‘본표제, 부표제’ 등과 더블린 코어의 ‘Title’을 이 클래스의 인스턴스로 정의함으로써 이종의 메타데이터에 대한 통합검색 방안을 제시하였다. 이 연구는 실제 검색시스템을 구현하여 성능을 시험하는 것이 목적도 아니었고 수행하지 않았으나 RDF를 이용한 메타데이터 통합모델의 제시에 그치지 않고 통합검색의 문제를 명확하게 다루었다는 점에서 큰 의미가 있다고 생각된다.

VI. 메타데이터 레지스트리를 통한 데이터 요소 표준화

RDF를 통한 메타데이터 포맷의 통합이 개별 포맷의 요소정의 및 기술을 인정하면서 호환성을 추구한다면 메타데이터 레지스트리는 유사 영역(domain)에서만이라도 가능하면 표준적인 스키마를 사용하여 데이터의 교환과 재활용을 위한 공통 접근방식을 취하고자 하는 노력이다. 따라서 RDF와 같은 기술(記述)의 틀은 이미 만들어진 이형의 메타데이터를 자신의 기술방식을 유지하면서 통합하기 위한 접근방식인 반면 메타데이터 레지스트리는 새로운 포맷의 생성 시 참조하여 중복 및 이형발생을 방지하고자 하는 예방적 접근방식인 동시에 기존의 데이터도 '표준'에의 접근을 용이하게 하는 방식이다.

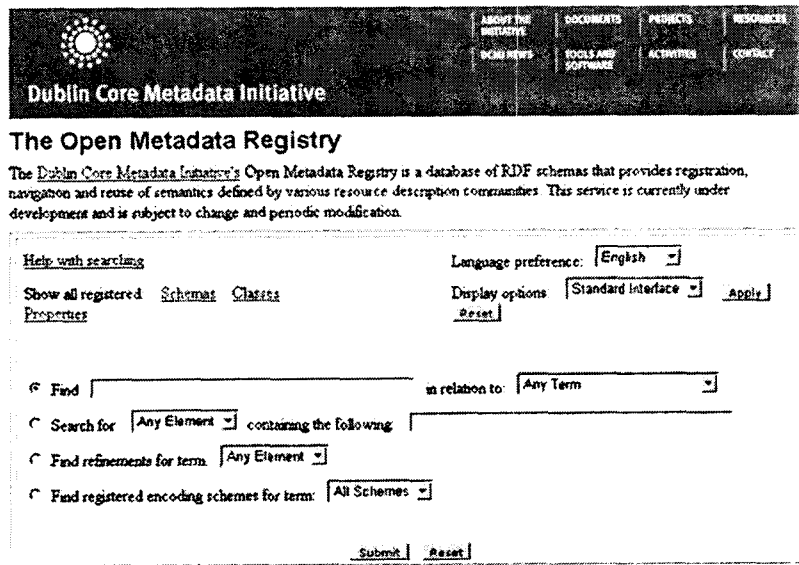
메타데이터 레지스트리는 사실 ISO/IEC 11179에 의하여 권장되는 공유 데이터 사전의 역사를 추적할 수 있는 스키마의 데이터베이스이며 등록절차이고, 근본적으로 메타데이터 레지스트리는 용어의 색인을 제공한다. 이러한 움직임은 영역(domain)과 표준화 커뮤니티 그리고 지식관리 커뮤니티에서 시작되었다 (Heery & Wagner, 2002).

메타데이터 레지스트리가 실현될 수 있는 방안은 예를 들면 구현자들이 생산한 스키마의 용어와 정의를 서로 연계하고 지역 스키마 관리인에 의하여 로컬에 저장하거나 또는 레지스트리가 다양한 메타데이터 스키마를 구현자들에게서 수집할 수 있다. 레지스트리는 특정 영역 또는 사용하는 커뮤니티에 적합한 스키마를 색인하여 유지함으로써 또는 한 눈에 복수 스키마를 접근할 수 있도록 용어의 네비게이션을 단순화함으로써 이용자들에게는 부가가치를 제공한다. 이 방식의 중요한 장점은 이용자가 새로운 용어를 생성하지 않고 현재 사용하는 용어의 재활용을 증가시킨다.

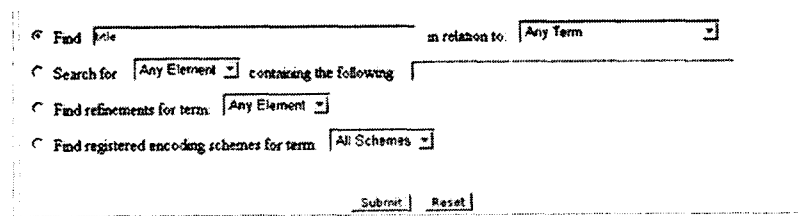
메타데이터 레지스트리는 등록기관이 구축되어야 의미가 있으며 상위설계에서 검증까지 전 단계가 일관되고 표준화되고 정형화되어야 동적(動的)인 현실세계의 모습을 표준화된 데이터 요소 형태로 데이터 레지스트리에 저장할 수 있다. 저장된 데이터 요소는 단계를 거치면서 기록된 데이터 요소(recorded data element), 승인된 데이터 요소(certified data element), 표준화된 데이터 요소(standardized data element)의 상태를 거치면서 동적으로 운영된다. 이는 실시간으로 수정되고 데이터를 필요로 하는 소비자들과 상호교류되어야 하므로 이를 지원하기 위한 자동화된 도구가 제공되어야 한다. 즉, 데이터 수집 과정에서부터 데이터 관리자는 정형화된 양식을 정보 제공자에게 제시할 수 있어야 하며 수집된 데이터를 해석하여 표준화된 데이터 요소와 속성들을 추출하여 저장할 수 있는 체계와 도구가 구축되어야 한다 (백두권 외, 1999).

메타데이터 레지스트리의 예는 외국에서 찾아볼 수 있으나 국내의 것은 문헌조사 결과 발견할 수 없었다. 외국의 사례 중 문헌정보학 분야의 레지스트리는 DCMI Registry Working Group의 Open Metadata Registry가 있다 (Wagner & Heery, 2001). 원래 이 레

지스트리의 목적은 DCMI 어휘에 대한 정보의 전거원을 제공하기 위하여 DCMI에 의하여 정의된 어휘의 등록, 발견 및 어의 네비게이션이 가능하도록 하여 DCMI의 활동을 지원하는 것이었으나, 그 관심이 더블린 코어 사용을 촉진하고 DVMI 어휘의 변경과 진화 관리를 지원하는 방향으로 바뀌었다.



〈그림 6〉 표준 이용자 인터페이스



Item(s) Found: 2

Item # 1	http://purl.org/dc/elements/1.1/title
Type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Property
Label	Title
Is Defined By	http://purl.org/dc/elements/1.1/
Definition	A name given to the resource.
Comment	Typically, a Title will be a name by which the resource is formally known
Item # 2	http://purl.org/dc/terms/alternative
Type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Property
Label	Alternative
Is Defined By	http://purl.org/dc/terms/
Refines	http://purl.org/dc/elements/1.1/title

〈그림 7〉 하이퍼링크된 검색결과

Open Metadata Registry의 궁극적인 목적은 레지스트리 응용프로그램을 위한 포괄적인 레지스트리 틀의 개발에 있었다. 이는 거듭되는 프로토타입을 구축 중이며 DCMI뿐만 아니라 다른 스키마도 포함하도록 확장성 있는 틀의 개발제공이라는 디자인 원칙을 가지고 진행하고 있다.

메타데이터의 통합 또는 통합검색이라는 측면에서 메타데이터 레지스트리는 기 개발된 메타데이터 표준을 이종의 표준과 통합하는 작업에 도움은 되나 직접적인 구조나 틀을 제공하지는 않는다. 또한, 다수의 메타데이터 포맷을 통합검색하는 방법을 제시하는 역할을 기대하기는 어려울 것으로 판단된다. 그러나 기 개발된 표준의 변이를 기록하고 관리하는 역할을 통하여 유사 영역의 메타데이터 생성에는 '전례'를 보여줌으로써 바람직하지 않은 다양성을 예방하는 역할이 있을 것이다. RDF와 관계는 우리가 앞서 언급한 내용 중 메타데이터를 통합하는 틀보다는 어의적 연계관계를 정의한 온톨로지와 유사한 특성을 가지고 있다.

Ⅶ. 결 론

다양한 메타데이터 표준이 나름대로의 필요성에 의하여 생성되었고 지금은 이들의 통합검색 수단으로 통합할 방법을 모색하고 있다. 이 글에서는 분산된 메타데이터 표준을 재통합하려는 노력이 과연 옳은가에 대한 판단을 내리지 않는다.

아마도 메타데이터 구현자들과 이용자 커뮤니티 간의 이해 상충의 결과이며 정보를 조직하는 목적이 정보를 필요로 하는 가장 많은 이용자에게 적절한 시기에 적합한 정보를 제공하는 것이라면 이중 메타데이터의 통합과 이들에 대한 통합검색은 필수적인 것이다.

이 글에서는 이중 메타데이터 간의 통합 및 통합검색의 도구로서 RDF와 메타데이터 레지스트리의 잠재성을 알아보고 이 둘의 관계를 알아보고자 하였다.

앞서 살펴 본 바에 의하면 RDF와 메타데이터 레지스트리는 동일한 목적을 위한 서로 다른 접근방법이 아닌 상호보완적인 관계에 있는 것으로 판단된다. 특히 메타데이터 레지스트리는 RDF의 구조적 결함이라고 할 수 있는 어의적 정의를 구성하는데 필수적 편이 장치이다: 첫째, 이는 온톨로지를 구성하는데 필요한 동일개념의 동의어 등의 용어간의 관계를 정의해 주며, 둘째, 이중 메타데이터 간의 공통 기술요소를 규명하는 목적으로도 사용될 수 있을 것이다.

그러나 메타데이터 레지스트리는 구현자 또는 이용자의 참여가 필요조건이므로 이에 대한 필요성을 바탕으로 관련 영역에서의 활발한 움직임을 기대해 본다. 메타데이터 레지스트리와 아울러 RDF는 특히 문헌유사객체를 위한 이중 메타데이터 통합의 기본 방향이다. 이와 관련하여 메타데이터의 통합뿐만 아니라 이들에 대한 통합검색 방안과 그들의

검색성능에 대한 연구가 진행되어야 할 것이다.

참고문헌

- 김이겸 (1999). RDF를 이용한 메타데이터 통합에 관한 연구 (박사학위논문, 연세대학교 대학원 문헌정보학과)
- 김이란 (2001). 온톨로지의 의미정보를 이용한 RDF 스키마 생성에 관한 연구. (석사학위 논문, 연세대학교 대학원 문헌정보학과)
- 김태수 (1999). 네트워크자원과 메타데이터. 정보관리연구, 30, pp. 1-26.
- 김태수, 김이겸 (1998). 메타데이터를 이용한 도서종합목록 구축에 관한 연구. 서울 : 첨단 학술정보센터.
- 김태수, 김이겸 (1997). 디지털 정보표현을 위한 메타데이터 표준개발에 관한 연구. 서울 : 첨단학술정보센터.
- 문헌정보처리연구회 (1998). 메타데이터의 형식과 구조. 서울 : 문헌정보처리연구회.
- 백두권 외 (1999). 최종연구보고서: 메타데이터 레지스트리, 데이터 요소 및 디지털 도서관 표준화 연구. [대전] : 한국전자통신연구원.
- Cathro. W. (1997). Metadata: an overview. [cited 2002. 12. 28].
(<http://www.nla.gov.au/nla/staffpaper/cathro3.html>).
- Chanldler. A., Foley, D. and Hafez, A.M. (2000). Mapping and converting essential Federal Geographic Data Committee (FGDC) metadata into MARC21 and Dublin Core: towards an alternative to the FGDC clearinghouse, D-Library Magazine, 6(1). [cited 2002. 12. 29].
(<http://www.dlib.org/january00/chandler/01chandler.html>)
- Cleverdon, C.W. (1962). Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield : College of Aeronautics.
- Day, M. (2001). Metadata in a nutshell. [cited 2002. 12. 28].
(<http://www.ukoln.ac.uk/metadata/publications/nutshell/>)
- DCMI (2003). DCMI Frequently Asked Questions (FAQ) : was Dublin Core metadata designed to be used only to described digital and Web-based resources? [cited 2003. 4. 25] (<http://dublincore.org/resources/faq/>)
- Dempsey, L. & Heery, R. (1998). Metadata: a current view of practice and issues. Journal of Documentation, 54(2), pp. 145-172.

- Diffuse (2002). Museum information standards. [cited 2003. 1. 5]
 (<http://www.difuse.org/museums.html>)
- Dooley, J. M. (Ed.) (1998). Encoded archival description : context, theory, and case studies. Chicago : Society of American Archivists.
- Duval, E. et al. (2002). Metadata principles and practicalities. D-Lib Magazine, 8(4).
 [cited 2003. 1. 14] (<http://www.dlib.org/dlib/april02/weibel/04weibel.html>)
- Heery, R. & Wagner, H. (2002). A metadata registry for the semantic web. [cited 2002. 11. 4] (<http://www.dlib.org/dlib/may02/wagner/05wagner.html>)
- Hickey. T. (1998). CORC - Cooperative Online Resource Catalog. D-Lib Magazine, 8(5).
 [cited 2002. 12. 28] (<http://www.oclc.org/research/publications/arr/1998/hickey/corc/htm>)
- FOLDOC : Free On-Line Dictionary of Computing (199?) [cited 2003. 7. 13]
 (<http://foldoc.doc.ic.ac.uk/foldoc/foldoc.cgi?Metadata>)
- Johnson, B. C. (2001). XML and MARC : which is "right"? Cataloging and Classification Quarterly, 32(1), pp. 81-90.
- Kokkkelink, S. & Schwanzl, R. (2002). Expressing qualified Dublin Core in RDF/XML.
 [cited 2003. 5. 5] (<http://dublincore.org/documents/2002/04/14/dcq-rdf-xml/>)
- Library of Congress (2001). Dublin Core/MARC/GILS crosswalk. [cited 2002. 12. 1]
 (<http://www.loc.gov/marc/dccross.htm>)
- Library of Congress (2001a). MARC to Dublin Core crosswalk. [cited 2002. 12. 1]
 (<http://www.loc.gov/marc/marc2dc.htm>)
- Lynch, C. et al. (1995). The nature of the NIDR challenge. (CNI White Paper: Chapter 1).[cited 2002. 12. 31] (<http://www.cni.org/projects/nidr/chapter-1.html>)
- McCallum, S. (2000). Extending MARC for bibliographic control in the web environment: challenges and alternatives. [cited 2002. 11. 27]
 (http://lcweb.loc.gov/catdir/bibcontrol/mccallum_paper.html)
- Miller, D. R. (2000). XML : libraries' strategic opportunity. Library Journal Net Connect, Summer. [cited 2002. 4. 22] (<http://xmlmarc.stanford.edu/LJ>)
- Pitti, D. V. (1998). Encoded archival description : the development of an encoding standard for archival finding aids. In J.M. Dooley (Ed.), Encoded archival description: context, theory, and case studies. Chicago : Society of American Archivists.
- SAS (2003). Data Warehousing Dictionary. [cited 2003. 7. 10]
 (<http://support.sas.com/md/warehousing/glossary.html>)
- Sherman, R. P. (1997). Metadata: the missing link. [cited 2003. 7. 5]

- (<http://www.dbmsmag.com/9708d16.html>)
- Society of American Archivists (1999). EAD application guidelines for version 1.0 : EAD crosswalks. [cited 2002. 12. 1] (<http://www.loc.gov/ead/agappb.html>)
- Spatial Data & Visualization Center(SDVC), U. of Wyoming (2002). Metadata education : what is metadata? [cited 2003. 4. 5] (<http://www.sdvc.uwyo.edu/metadata/what.html>)
- Sparck Jones, K. (1981). The Cranfield test. In K. Sparck Jones (Ed.), *Information retrieval experiment* (pp. 256-284). London : Butterworths.
- St. Pierre, M. & LaPlant, W.P., Jr. (1998). Issues in crosswaling content metadata standards. [cited 2003. 1. 14] (<http://www.niso.org/press/whitepapers/crosswalk.html>)
- TEI guidelines for electronic text encoding and interchange : historical background. [n.d.]. [cited 2003. 1. 14] (<http://etext.lib.virginia.edu/bin/tei-tocs?div=DIV2&id=ABTEI>)
- UKOLN Metadata Group (1998). A review of metadata : a survey of current resource description formats. [cited 2002. 12. 28] (http://www.ukoln.ac.uk/metadata/desire/overview/rev_toc.htm)
- W3C RDF Core Working Group (2003). RDF Primer. [cited 2003. 4. 12] (<http://www.w3org/TR/2003/WD-rdf-primer-2003-123>) (번역본: RDF 입문서. 심 경 번역. 발간예정)
- Wagner, H. & Heery, R. (2001). DCMI Open Metadata Registry purpose and scope : phase 1. [cited 2003. 4. 6] (<http://dublincore.org/groups/registry/scope-20011108.shtml>)
- Weibel, S. (1995). Metadata : the foundations of resource description. *D-Lib Magazine*, July '95. [cited 2003. 1. 14] (<http://www.dlib.org/dlib/july95/07weibel.html>)
- Weibel, S. & Iannella, R. & Cathro, W. (1997). The 4th Dublin Core Metadata Workshop Report : DC-4, March 3-5, 1997, National Library of Australia, Canberra. [cited 2002. 12. 31] (<http://www.dlib.org/dlib/june97/metadata/06weibel.html>)