

# Dublin Core 형식의 개선점에 관한 연구

## A study on the improvements in Dublin Core

강 유 정, 이화여자대학교 문헌정보학과

Yu-Jeong Kang, 972LSG10@mm.ewha.ac.kr  
Dept. of Lib. and Inf. Sci., Ewha Womans Univ.

현재 인터넷자원은 분산된 네비게이션환경에서 배포되고 있다. 이러한 분산된 정보자원을 통합할 수 있는 표준형식으로 Dublin Core가 성장하기 위해서는 기술내용뿐만 아니라, 기술방법을 정의하여 작성된 레코드의 일관성을 확보하여야 한다. 그리고 더 나아가 다언어문화도 자료의 유형에 상관없이 통합 기술할 수 있는 보다 세련된 메타데이터 형식으로 발전시켜야 한다.

### 1. 서론

현재 인터넷 정보검색이 지닌 한계는 다음과 같다. 첫째, Altavista, Webcrawler 등과 같은 탐색엔진의 전문색인을 통한 검색결과 신뢰도가 낮다는 점이 있다. 둘째, 공동의 표준이 적용되고 있지 않으므로 여러 사이트에 분산된 정보자원이 통합되지 못하고 있다는 점이 지적되고 있다. 게다가 관련 정보자원간의 연계 또한 사실상 이루어지지 않고 있으므로 관련 정보자원을 검색하기 위해서는 해당 자료의 목록을 재조사해야 하는 번거로움이 있다.

이에 이러한 문제점들의 해결에 노력하고자 메타데이터 형식이 각 주제분야와 서지기관에서 개발되었다. 메타데이터 형식 또한 분산환경에 따라 서로 상이하게 출현하였는데 이는 각 주제 분야의 이용자 수준과 요구 그리고 데이터요소와 수준이 각각 다르기 때문에 생긴 현상이다. 그러나 이렇게 각각 다른 형식으로 메타데이터 형식들이 개발되었기 때문에 자원 기술의 일관성과 데이터의 호환성이 낮으며 이런 문제는 전술한 인터넷자원의 검색에서 분산된 정보자원의 통합문제가 아직도 해결되지 못하고 있다는 것을 암시해 준다. 따라서 분산된 정보자원의 통합을 위한 표준으로 또는 연계를 제공할 수 있는 데이터 상호교환형식으로 Dublin Core Metadata Element Set (이하 Dublin Core라 한다)가 생성되었다.

그러나 이런 Dublin Core가 그 생성동기에 비해 그리고 서지레코드로 활약하기에는 아직 미

흡한 점이 많으므로 본고에서는 Dublin Core 형식의 합리적인 전개안을 도출하기 위하여, 이 형식이 갖고 있는 제반적인 문제점과 연구과제를 제시하고자 한다.

### 2. Dublin Core의 문제점

#### 2.1 기술원칙

Dublin Core는 목록전문가가 아니더라도 자료의 저자나 배포자가 웹 자원을 직접 기술할 수 있도록 하는 단순구조를 주요 특징으로 한다. 이 형식은 15개 핵심요소로 구성되며 기본 기술원칙으로는 오브젝트의 본질적인 속성의 기술, 확장성, 선택성, 반복성, 수정가능성 그리고 구문독립성이 있다(Weibel 1995). 이런 기술원칙 상의 문제점을 제시하면 다음과 같다.

첫째, 기술원칙 중 기본기술원칙 상에 문제가 있는데, 예를 들면 확장성, 수정가능성, 구문독립성이 이에 해당된다. 먼저, 확장성 메카니즘을 통해 서지기관이나 작성자가 필요시 원하는 요소를 추가할 수 있다. 그러나 이런 특성 때문에 Dublin Core 형식이 무한정 확장되거나 서지기관마다 형식이 다르게 전개될 수도 있으므로 표준화의 저해요인이 된다. 그리고 수정가능성은 서로 다른 주제분야의 요구를 만족시켜주기 위하여, 한정어 값으로 요소의 정의를 수정하여 사용할 수 있다는 것이다. 그러나 이 원칙은 요소정의의 일관성을 저해하는 요인이 된다. 구문독립성이란, Dublin Core 형

식을 각 주제분야와 응용 프로그램의 범위에서 각자 사용할 수 있도록 특정 구문으로 제한하지 않는 것을 의미한다. 그러나 이런 원칙은 같은 Dublin Core 형식으로 기술된 레코드간의 호환성에 문제를 발생시킨다. 따라서 이런 원칙들을 통하여 기술에 융통성을 제공하기보다는 형식의 표준화를 통하여 기술에 일관성을 제공하는 것이 더 바람직할 것이다.

둘째, 일정한 순서 없이 기술하므로 이 형식 내의 관련 데이터요소가 분산되어 레코드의 내용식별에 용이하지 못하다. 따라서 관련 있는 데이터요소들을 모아주는 방향으로 기술순서를 정해야 한다. 예를 들면, 제작자 관련사항인 제작자요소(Creator)와 기타제작자요소(Contributor)를, 발행 관련사항인 발행처요소(Publisher)와 발행일자요소(Date)를 분산하여 기술하지 않고 모아주는 것이 레코드의 내용식별에 효과적일 것이다.

셋째, 기술의 일관성이 결여되어 있다. 따라서 전거제어와 기술방법의 정의를 통해 기술의 일관성과 검색결과의 정확성을 높힐 수 있게 하여야 한다.

넷째, 기술단위가 결정되지 않았다. 따라서 개별자원단위로 기술할 것인지 또는 웹 사이트와 같은 장서단위로 기술할 것인지를 결정해야 한다.

## 2.2 인터넷자원의 검색도구적 측면

Dublin Core(한정어도 포함함)와 같은 탐색의 범주에 해당되는 메타데이터들과 비교해 본 결과 선반적으로 정보자원의 식별과 확인정보를 제공하는 기술요소가 부족하며 인터넷자원의 특성을 잘 반영하지 못하고 있다.

비교내용을 구체적으로 제시하면 다음과 같다.

첫째, 이용조건이나 이용적합성 정보를 제공하는 요소가 부족하다. 네트워크자원은 주로 위치정보를 통해서 액세스한다. 그러나 인터넷자원을 액세스하기 위해서는 주소뿐만 아니라, 액세스방법, 액세스제한요소(예를 들면, 패스워드나 로그인 등의 확인정보)와 같은 구체적인 내용이 필요하다. 그리고 이용자가 이용적합성을 판단할 수 있는 정보를 제공할 수 있어야 한다. 이용적합성이란, 이용자의 하드웨어와 소프트웨어, 네트워크환경, 자원의 내용과 포맷, 액세스방법이 복합적으로 관련 있는 성질로, 실제 특정자원을 이용하는 능력과는 구분된다(김태수, 김이경 1997). 따라서 이런 이용적합성을 이용자가 판단할 수 있도록 하위요소가

전개되어야 한다.

둘째, 관련자원에 연결정보를 제공하는 능력이 부족하다. 비교대상 형식은 주로 참조정보라는 요소로 제시하고 있으나 Dublin Core 관련자료요소(Relation)의 한정어 값에 대한 여러 견해가 통일되어 있지 않은 상태이다. 여기에서 고려해야 할 사항은 1:1의 원칙을 준수해야 하기 때문에 관련자원에 대해 단순기술하고 연결정보를 제공할 수 있는 정도로 전개되어야 한다. 뿐만 아니라, 미리 사이트와의 연결정보 제공을 고려해야 한다.

셋째, 정보자원의 최신성과 리뷰, 갱신 또는 철회에 관한 내용을 기술하는 요소가 부족하다. 인터넷자원에서 중요한 사항으로 발행일자 정보를 들 수 있다. 인쇄문헌에서는 날짜사항을 주로 년단위로 처리하나 전자문헌은 수정과 변경이 빈번하기 때문에 일단위나 시분초단위로 처리하게 된다. 또한 정보의 최신성이 중요하며 데이터에 대한 리뷰와 갱신에 관한 정보가 정보자원의 신뢰성 판단과 식별에 중요한 인이 되기도 하므로 Dublin Core 형식의 발행일자요소에서는 이러한 사항들을 확인하고 식별할 수 있도록 하위요소를 전개해야 한다.

넷째, 인쇄문헌을 디지털화 하여 웹에 올려 놓은 학술잡지들이 많음에도 불구하고 Dublin Core에는 이런 연속간행물의 권호나 PDF인 경우 페이지정보를 기술할 수 있는 요소가 실정되지 않았다. 그리고 인터넷상에는 주로 공식적인 출판보다 배포의 성격을 많이 갖는다. 따라서 배포자(Disseminator)와 발행처요소를 식별할 수 있도록 해야 한다.

다섯째, 주로 텍스트정보 위주로 하위요소가 고려되었다. 그러나 한 형식에서 모든 유형의 자원을 포괄할 수 있도록 하위요소가 전개되어야 한다. 예를 들면, 이미지 데이터 검색을 위해서 내용범위요소(Coverage)의 한정어 값을 좌표 값에 기반을 둔 형식으로 제안한 예가 있다. 뿐만 아니라, 슬라이드 자료인 경우에는 슬라이드 총수를 그리고 동화상이나 소리데이터인 경우, 총 재생시간 등과 같은 데이터의 특성을 반영하여 하위요소가 전개되어야 한다.

이외에도 데이터 전송시간, 파일의 크기, 파일압축정보, 기술적 선행조건 등을 기술하여 이용자로 하여금 실제 자원에 액세스여부를 미리 결정할 수 있게 하여야 한다.

전술한 내용을 반영하여 Dublin Core 한정어를 설계할 때 고려해야 할 점은 형식이 무한정 확장될 수는 없으므로 "상세"수준의 제한범위를 정하는 일이다.

### 2.3 서지레코드적인 측면

전 세계적으로 이 형식을 이용하여 서지레코드를 작성하고자 하는 시도가 이루어지고 있다. 그러나 서지레코드로 완전하게 기능하기에는 미비한 점이 많으므로 다음의 내용을 보완할 필요가 있다(김태수, 김이경 1997).

첫째, 탐색기능의 측면에서 Dublin Core는 웹 자원의 탐색에 가장 중요한 요소만을 데이터로 수용하나 이런 요소기술에 일관된 원칙이 적용되지 못하고 있으므로 부적합 문헌이 다수 검색된다. 따라서 제어어휘를 사용하여 탐색기능을 지원할 필요가 있다.

둘째, 확인의 기능 측면에서 Dublin Core는 탐색의 기능에 주안점을 둔 형식이지만 특정자원의 확인과 관련자원의 식별에는 그 기능이 제한된 형식이다. 예를 들면, 전자자원의 버전은 자원의 식별과 확인에 중요 근거가 되나 이를 필수적으로 기술할 요소가 없다.

셋째, 자원선정의 기능 측면에서 이용자가 다양한 자원의 유용성을 평가할 수 있는 정보를 제공하는 데이터요소가 없다. 물론 설명요소(Description)나 내용범위요소(Coverage)에서 언급할 수 있으나 유용하지는 않다.

넷째, 자원확보의 기능 측면에서, 식별자요소(Identifier)에서 네트워크자원의 위치정보를 제공하도록 하고 있으나 URL과 같은 주소는 쉽게 변한다. 따라서 정확한 주소를 제공하여 점차 안정된 식별수단으로 기능할 수 있도록 해야 한다.

전술한 문제점들 때문에 ALA(1988)에서는 서지데이터베이스에 포함되지 않고 별도로 유지되어야 한다고 제안하고 있고 이외의 문제점으로 다음의 내용을 지적하였다.

첫째, 메타데이터는 목록의 대체수단이 아니라 웹 자원의 탐색을 위해 개발된 것이다.

둘째, 전거파일이나 전거제어수단을 적용하지 않으므로 도서관목록으로 사용하는 것은 부적절하다.

셋째, Dublin Core의 데이터를 MARC으로 변환할 수 있지만 일관성과 표준화라는 점에서 목록규칙이 규정하고 있는 조건에 부적합하다.

넷째, 데이터 등록과 유지관리에 책임 있는 기관의 지원을 필요로 한다.

따라서 서지레코드로 기능하게 하기 위해서는 전술한 내용이 개선되어야 할 것이다.

### 2.4 다언어문화의 공존 문제

정보의 세계가 지구촌화 되면서, 지배적인

언어인 영어권의 정보자원뿐만 아니라, 비지배적인 언어권에서 작성된 정보자원에 대한 액세스의 중요성이 증가하고 있다. 그리고 여러 언어권에서 자국의 기술구조에 적합한 형식으로 버전을 달리하여 Dublin Core 형식을 조금 다르게 전개하고 있는 실정이다.

일반적으로 상이한 언어권에서 생성된 정보의 교환을 위하여 주로 3가지 방식이 사용되고 있다. 하나는 공통의 언어표현법이나 의미를 지시하는 방법이고 또 하나는 공통의 프로토콜을 사용하는 것이고 나머지 하나는 번역시스템을 통해 1:1로 번역하여 데이터를 교환하는 것이다. MARC과 같은 메타데이터에서는 두 번째와 세 번째 방법을 이용할 수밖에 없다. 그러나 세 번째 방법은 특정 두 언어간 개별작용이 있어야 하므로 다언어 시스템은 채용하기가 어렵고 두 번째 방법은 세 번째나 첫 번째 방법이 선행되어야 한다는 점의 문제가 있다.

첫 번째 방법에서 공통의 언어표현을 사용하면 다음과 같은 제약이 따른다.

첫째, 때론 공통의 의미기호로 표현되지 못할 경우가 있다.

둘째, 한 항목이 다수의 의미로 번역될 수 있다. 정보검색에 있어서 다언어문제에 대한 대부분의 견해가 다언어간의 유니코드(unicode)를 변환하면 가능하다고 생각한다. 그러나 단순한 코드변환은 요소 값이 담고 있는 의미까지 표현하지는 못하므로 부적합문헌이 검색되는 예가 허다하다. 이런 문제는 특히 한자문화권에 있는 국가에서 많이 발생한다. 예를 들어, 'The descriptive changes of catalog'라는 표제의 기사가 있다고 가정해 보자. 이 표제는 한글로 번역하면 '목록의 기술변화'가 된다. 그러나 한글로 번역된 전술된 문장은 2가지 의미로 해석된다. 하나는 목록의 記述 변화이며 또 다른 의미는 목록의 技術변화가 된다. 즉, 전자의 변화는 문헌정보학적 관점에서 記述의 변화를 의미하며, 후자는 技術工學적 변화를 의미한다.

셋째, 메타데이터 생성자가 한 언어뿐만 아니라, 새로운 공통의 의미기호를 마스터해야 하는 부담을 갖게 된다.

뿐만 아니라, 정보검색자가 기술양식과 타겟 언어를 반드시 알고 있어야만 한다는 제한점과 검색범위가 세계적이지 못하고 공통의 기술양식을 적용하는 데 방법상의 문제가 각 언어마다 있다는 문제가 있다.

이런 제약에도 불구하고 다언어간의 정보교환을 위해서는 공통의 의미기호와 공통의 표현양식이 필요하다. 뿐만 아니라, 다언어문화가

공존하고 있는 언어권에서는 특별히 고려해야 할 문제가 있다. 일 예로, 우리 나라의 경우에는 한자가 여러 가지 의미로 해석될 수 있기 때문에 단순한 번역이 데이터 해석의 기초가 되어서는 안된다. 따라서 이런 문제를 Dublin Core 한정어의 설계에 반영하여 데이터 교환과 검색의 정확도를 높힐 수 있도록 하여야 한다.

따라서 만약 Dublin Core에 다언어문화를 고려하여 형식을 설계하고자 한다면, 다음과 같은 내용을 반영할 수 있을 것이다. 먼저 공통의 의미기호로는 당연히 영어가 될 수밖에 없고 이 영어가 피벗 언어의 역할을 하게 될 것이다. 그리고 원 언어(original language)와 영어로 표현한 의미(meaning)를 한정어 값으로 설계할 수 있을 것이다. 예를 들면 다음과 같다.

```
<meta name="DC.title" type="original" lang="ko" content="目錄의 記述變化"/>
<meta name="DC.title" type="meaning" lang="en" content="The descriptive changes of catalog"/>
```

따라서 다언어문화를 반영할 수 있는 한정어를 설계하여 Dublin Core 형식을 통합형식으로 발전시켜야 한다.

### 2.5 상호운용의 문제

상호운용성을 위하여 Dublin Core 형식은 첫째, 한정어를 제한하여야 하며 이를 위해서는 각 어플리케이션에서 가능한 DDC, MeSH, LCSH 등과 같은 외부표준을 따르는 것이 바람직할 것이다.

둘째, 상호운용성이 최소 수준 즉, 데이터 상호교환과 탐색수준 정도에서 반드시 이루어져야 한다. 먼저 데이터교환을 위해서 공통의 의미를 갖고 공통의 표현구분과 구조를 공유해야 한다. 그리고 탐색시 상호운용성을 위해서는 두 어플리케이션간의 공통의 의미론과 질의 및 응답을 위한 프로토콜의 공유가 최소한 필요하다. 따라서 전술된 측면에서 표준화가 이루어져야 한다.

### 3. 결론 및 제언

어플리케이션 상호간 일관성 있게 전개될 수 있는 좀 더 상세하고 표현력이 풍부한 버전을 표준화함으로써, 메타데이터 교환과 탐색을 위한 학제적 상호운용성의 범위를 확대해야 할 것이다. 무엇보다도 요소값의 전거제어를 통해

지식구조를 효과적으로 유지할 수 있도록 해야 한다. 앞으로 더 논의되어야 구체적인 영역은 다음과 같다.

첫째, 표준화를 위한 노력이다. 최초로 공식적인 요소정의란 한 문헌이 RFC 2413인데 이 문헌에 약간의 수정을 가할 필요가 있다. 그 수정의 범위는 먼저, 요소정의를 명료하게 하고 일관성 있게 전개해야 하며 ISO 11179로 알려진 메타데이터 요소를 위한 표준기술템플릿에 따라 Dublin Core 명세를 구성하는 것이다. ISO 11179는 일관성 있는 방법으로 데이터 요소의 의미론을 공식적으로 표현하기 위한 국제표준이다(Weibel 1999).

둘째, HTML로 Dublin Core 형식을 인코딩하는 문제가 있다. 이미 제시된 규약이 비공식적인 것이기 때문에 이것을 대신할 수 있는 공식적인 명세가 필요하다.

셋째, Dublin Core 요소를 상세하고 일관성 있게 할 수 있는 메카니즘을 기초적인 데이터 모형에 제공할 수 있어야 한다. 이를 위하여 먼저, 특정주제분야의 통제어휘나 분류체계를 사용하면 기술적인 정확도를 높힐 수 있다. 그리고 요소 값을 분명하게 하기 위한 공식적인 인코딩 표준을 확보해야 한다. 그 대표적인 예로 발행일자요소를 들 수 있다. 또한 공식적인 하위구조를 정의해야 한다.

넷째, 자원유형별로 형식을 다르게 전개해 나갈 것이 아니라, 한 형식에 통합 기술할 수 있도록 연구되어야 한다.

다섯째, Dublin Core 형식에 다언어문화를 반영하여 각국에서 다른 버전으로 설계된 형식을 통합할 수 있어야 한다.

### 참고문헌

- 김태수, 김이경. 1997. 「디지털 정보표현을 위한 메타데이터 표준개발에 관한 연구」. [서울]: 첨단학술정보센터.
- ALA. *Committee on Cataloging Task Force on metadata and the cataloging rules*. 1988. [cited 1999.06.23]. <<http://www.ala.org/alcts.organization/ccs/cc.da/tf-tei2/>>.
- Weibel, S. 1995. "Metadata: the foundations of resource description." [cited 1999.06.24]. <<http://www.ala.org/alcts.org.dlib/july95/07weibel/>>.
- . 1999. "The state of the Dublin Core metadata initiative april 1999." [cited 1999.06.23]. <<http://www.dlib.org/dlib/april99/04weibel.html>>.