



네트워크 환경에서 대표적인 전자 자원의 기술 수단

연재순서

1. 메타데이터의 개요
2. DC(Dublin Core) 이번호
3. GILS(Government Information Locator Service)
4. IAFA Templates
5. MARC
6. PICS(Platform for Internet Content Selection)
7. RFC 1807
8. SOIF
9. TEI header
10. URC(Uniform Resource Characteristics)
11. Warwick Framework RDF(Resource Description Framework)
12. 메타데이터 향후방향

더블린 코어는 더블린 메타데이터 코어 요소 집합(Dublin Metadata Core Element Set, 이하 DC)을 일컫는 말로서, 1995년 미국 더블린에서 열린 1차 워크샵을 통해 정의되었다. DC는 네트워크 환경에서의 각종 전자 정보를 기술하는데 핵심적인 데이터 요소로 현재 진행 중인 메타데이터 관련 연구 작업의 기초가 되고 있다.

지금까지 DC에 관한 5차례 걸친 워크샵을 통해 요소 집합의 정의 및 수정·보완, 데이터를 기술하는데 있어서 다루어야 할 의미·구문론, 그리고 다양한 메타데이터와의 상호운용을 위한 통합 작업과 관련된 논의를 토대로 메타데이터의 표준화 작업이 추진되고 있다.

1. DC 워크샵

· DC-1

DC-1(1차 워크샵, 이하 동일 표현)은

1995년 3월 OCLC(Online Computer Library Center)와 NCSA(National Center for Supercomputing Applications)가 미국 더블린(Dublin)에서 개최하였다. DC-1에는 문헌정보학, 인문학, 지리학, 컴퓨터과학 등 관련 분야의 전문가와 전문 사서, 연구자들이 참석하여 정보 접근을 쉽게 할 수 있도록 네트워크상의 광범위한 주제 분야의 전자 정보를 간단하게 기술하고 접근하기 위한 13개의 메타데이터 요소 집합을 정의하였다. 여기에서 DC의 명칭이 유래되었다.

· DC-2

1996년 4월 영국의 워릭(Warwick)에서 열린 DC-2는 OCLC와 UKOLN(UK Office for Library and Information Networking)이 주관하여 개최되었다. DC-2에서는 웹기반 환경에서도 이용가능하고 입력 및 교환을 위한 구문

〈표 1〉 DC 워크샵 개요

	일시	장소	주요 내용
DC-1	1995년 3월	미국 더블린	· DC 13개 데이터 요소 집합 정의
DC-2	1996년 4월	영국 워릭	· DC 구문 제안 · 워릭 프레임워크 제안
DC-3	1996년 9월	미국 더블린	· 이미지 자료으로 확장 · 데이터 요소 추가(총 15개)
DC-4	1997년 3월	호주 캔버라	· 정형화된 한정어
DC-5	1997년 10월	핀란드 헬싱키	· 데이터요소 기술 사항 정의
DC-6	1998년 11월	미국 워싱턴	(개최 예정)

정보자원을 기술하는 메타데이터는 자원의 식별이나 이용 타당성 평가, 자원 유지를한 자원의 특징 파악 등을 위해 사용된다. 따라서 여러 가지 목적에 맞는 서로 다른 이용자 집단에서 다양한 메타데이터 표준을 개발하여 사용하고 있다. 더블린 코어(Dublin Core)는 다양한 메타데이터 표준 가운데 하나로서 네트워크상의 광범위한 자원을 기술하기 위한 간단하지만 효과적인 메타데이터 요소 집합이다.

이재진/ 한국데이터베이스진흥센터 정책연구과

(syntax)과 데이터 상호교환을 위해 다양한 메타데이터를 모으는 컨테이너 구조인 워릭 프레임워크(Warwick Framework)가 제안되었다.

· DC-3

DC-3은 1996년 9월 미국 더블린에서 OCLC와 CNI(Coalition for Networked Information)의 주관으로 개최되었다. DC-3에서는 DC가 기술하는 대상 자원을 텍스트가 아닌 자원, 즉 이미지 자원 등으로 확장하기 위해 데이터 요소의 개정을 추진하였고, 그 결과 2개 요소를 추

가하여 기본 데이터 요소를 15개로 확정하였다.

· DC-4

OCLC, DSTC(Distributed Systems Technology Centre), 그리고 NLA(National Library of Australia)가 주관하여 1997년 3월 호주 캔버라(Canberra)에서 열린 DC-4에서는 데이터요소의 하부구조를 기술하기 위한 정형화된 한정어와 다른 메타데이터로의 확장 방안, 그리고 데이터 요소의 의미와 기술 범위의 내용을 규정하였다.

· DC-5

가장 최근인 1997년 10월 핀란드의 헬싱키(Helsinki)에서 열린 DC-5는 OCLC와 National Library of Finland가 주관하였다. DC-5에서는 이전에 정의된 데이터 요소를 수정·보완하여 15개 데이터 요소의 상세 하부구조와 그 기술 사항의 정의가 마무리되었다.

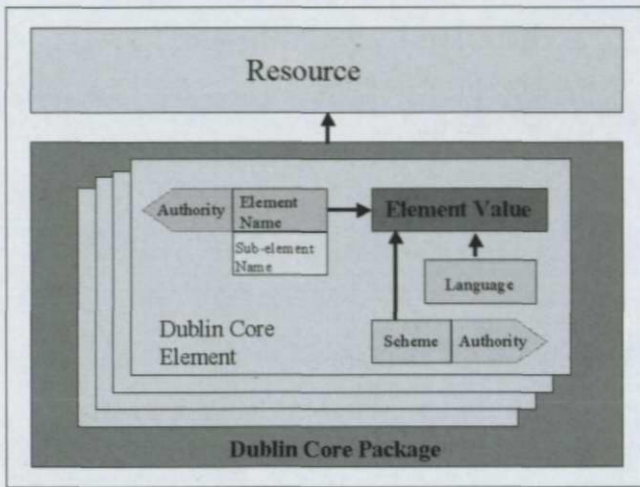
2. DC 개발 원칙

DC-1에서 처음 메타데이터 요소를 개발할 때 가장 염두에 둔 점은 포괄적이지만 최소의 요소 집합을 정의하는 것이었다. 그러나 최소의 요소 집합을 정의하되 이를 모든 이용자가 쉽게 이해할 수 있고 광범위한 주제 영역의 자원을 기술하기에 충분한 유연성을 가져야 하는 과제를 안게 되었다.

이러한 점을 수용하기 위해 자원의 본질적인 특성을 기술하는데 역점을 둔다는 고유성(Intrinsicality), 정의된 데이터 요소만으로는 나타내기 어려운 내용도 기술할 수 있도록 하는 확장성(Extensibility), 응용분야나 표현기법을 규정하지 않는 구문 독립성(Syntax-Independence), 모든 요소를 선택사항으로 정하는 선택성(Optionality), 모든 요소를 여러 번 사용할 수 있는 반복성(Repeatability), 그리

〈표 2〉 DC 메타데이터 요소

범주	데이터 요소	설명
내용 (Content)	Title	제작자(creator)나 발행자(publisher)가 사용한, 자원에 부여한 이름
	Subject	자원의 주제나 내용을 기술하는 단어나 구, 제어어집합·분류표 이용
	Description	DLO(document-like objects)는 초록, 시각자료는 내용 기술, 문장형식
	Source	현재의 자원의 출처가 되는 정보
	Language	자원의 내용을 기술한 언어. RFC 1766(언어식별태그) 코드 사용
	Relation	독립적으로 존재하는 자원간의 공식적인 관계 표현. 관련된 자원과 지시된 자원 기술 연결
	Coverage	자원에서 취급하고 있는 내용의 공간·시간적 특징. 공간적 특징은 지명이나 좌표를 사용한 물리적 지역(예, 위도, 경도), 시간적 특징은 시기를 나타내는 단어나 구(예, 신석기 시대)를 의미
지적 소유권 (Intellectual Property)	Creator	자원의 내용에 책임이 있는 인물이나 기관
	Publisher	현재의 형태로 제작한 자원의 제작 기관
	Contributor	Creator 요소에 기술된 인물·기관에 비해 부차적 성격의 인물·단체
	Rights	자원의 판권 관리사항 및 관련 정보의 제공서비스 연결 식별어
형식 (Instantiation)	Date	자원이 제작되거나 이용가능하게 된 날짜. YYYY, YYYY-MM-DD형태
	Type	홈페이지, 소셜, 시, 수필, 사진 같은 자원의 유형. 지정리스트 이용
	Format	자원의 데이터 포맷과 차원(크기, 지속기간). 지정리스트 이용
	Identifier	자원을 고유하게 식별하기 위한 문자열·기호. URL, URN, ISBN 등



(그림 1) DC 모델

(표 3) DC 한정어

데이터요소	Sub-element	Scheme
Title	· Main · Alternative	x
Creator	· PersonalName · PersonalName.Address · CorporateName · CorporateName.Address	LCNAF
Subject	x	Keyword(default), LCSH, MeSH, AAT, LCNAF, DDC, LCC, NLM, UDC
Description	x	Abstract(default), URL
Publisher	· PersonalName · PersonalName.Address · CorporateName · CorporateName.Address	x
Contributors	· PersonalName · PersonalName.Address · CorporateName · CorporateName.Address	LCNAF
Date	· Created · Issued · Accepted · Available · Acquired · DataGathered	ISO 8601(default) ANSI X3.30 IETF RFC 822
Type	· Valid	Controlled List
Format	x	IMT(MIME), DCPMT
Identifier	x	URL(default), URN, ISBN, ISSN, SICI, FPI
Source	x	free text(default), URL, URN, ISBN, ISSN
Language	x	IETF RFC 1766, Z39.53, ISO 639-1, ISO 639-2/B
Relation	· Creative · Mechanical · Version · Inclusion	free text(default), URL, URN, ISBN
Coverage	· Reference · PeriodName · PlaceName · t · x · y · z · Polygon · Line · 3d	· Spatial scheme free text(default), DMS, DD, OSGB, UTMXX, LCSH, TGN · Temporal scheme free text(default) LCSH, AAT, LSN, ISO 8601, ANSI X3.30
Rights	x	free text(default), URL, URN

고 선택적인 한정어를 사용하여 각 요소의 세부사항을 조정할 수 있도록 하는 수정가능성(Modifiability)을 개발 원칙으로 세우고 이를 전제로 데이터 요소 집합을 정의하였다.

3. DC 메타데이터 요소

DC는 DC-1에서 13개의 요소가 정의되었고, 이후 DC-3에서 2개의 요소(DESCRIPTION, RIGHTS)가 추가되어 15개의 요소로 정의되고 있다.

15개의 요소는 크게 자원의 내용과 관련된 요소와 지적 소유권 측면에서 본 자원과 관련된 요소, 그리고 자원의 형식과 관련된 요소의 3가지 범주로 구분된다. 15개의 요소는 다음 <표 2>와 같다.

4. DC 모델

호주 캔버라에서 열린 DC-4에서는 데이터의 의미적 내용을 기술하기 위한 논의가 이루어졌다. 이는 DC 표준의 구조 확장과 관련된 것으로서 이른바 최소화(minimalist)와 구조화(structuralist)의 대립으로 일컫는다. 최소화는 데이터의 의미적인 표현을 위해 최소한의 한정어를 정하고 그 하부구조를 설정해야 한다는 것이고, 구조화는 다양한 요구에 맞게 더욱 정확하게 요소를 표현해야 한다는 것이다.

이러한 논의 끝에 DC-4에서는 Language, Scheme, Sub-element의 3가지 한정어를 제안하였고 이를 캔버라 한정어(Canberra Qualifiers)라 한다. 각 데이터 요소는 요소값(내용) 외에 이들 한정어를 통해 의미적 표현을 할 수 있다.

한정어를 포함하여 자원의 메타데이터를 DC로 표현한 DC 모델을 도식화하면 <그림 1>과 같다.

자원에 대한 메타데이터를 DC로 표현할 때, 각 메타데이터 요소들의 집합을 DC 패키지라 한다. 다시말해 자원의 메타데이터는 DC 패키지로 표현할 수 있다. 각 요소는 요소값(element value)을 갖는데, 요소값은 한정어를 통해 그 의미를 확장한다. 즉, 요소명(element name)의 하부구조를 한정어 Sub-element로 표현하고, 이밖에 Language, Scheme과 같은 한정어를 사용한다.

한정어 Language는 자원 자체의 언어를 나타내는 데이터 요소 Language와는 달리 요소값을 표현하는 언어를 나타내준다. 이것은 점차 웹환경에서 사용되는 문서의 언어가 다중화되어감에 따라 자원의 언어와 자원을 기술하는 언어의 식별이 중요한 문제가 되고 있는데 따른 것이다.

한정어 Scheme은 주어진 요소의 내용을 해석하는데 기준이 되는 외부 체계를 나타내는 것으로, 분류체계나 코딩체계, 시소러스 등이 해당된다. 한정어 Sub-element는 요소의 내용을 더욱 분명하고 상세하게 설명하기 위한 것으로 요소명의 하부구조를 표현한다. 이들 한정어는 DC의 15개 데이터 요소 각각에 선택적으로 사용될 수 있으며, 특정 정보를 더욱 상세하고 의미있게 나타내어 시스템에서의 탐색이나 검색의 정확성을 높이는 데 기여한다.

DC-4에서 제안된 한정어 및 하부구조에 관한 논의는 DC-5에서 더욱 심층적으로 다루어져 각 데이터 요소마다 관련된

```
<META NAME= "DC.Title" CONTENT= "Song of the Open Road">
<META NAME= "DC.Creator" CONTENT= "Nash, Ogden">

<META
NAME= "DC.subject"
CONTENT= "(SCHEME=LCSH) (LANG=EN)
Computer Cataloging of Network Resources">
<META
NAME="DC.coverage.periodName"
SCHEME="historic"
CONTENT="Ming Dynasty">
```

<예 1> HTML의 DC 표현 - META 태그

```
<LINK REL= SCHEMA.dc HREF="http://purl.org/metadata/dublin_core">
```

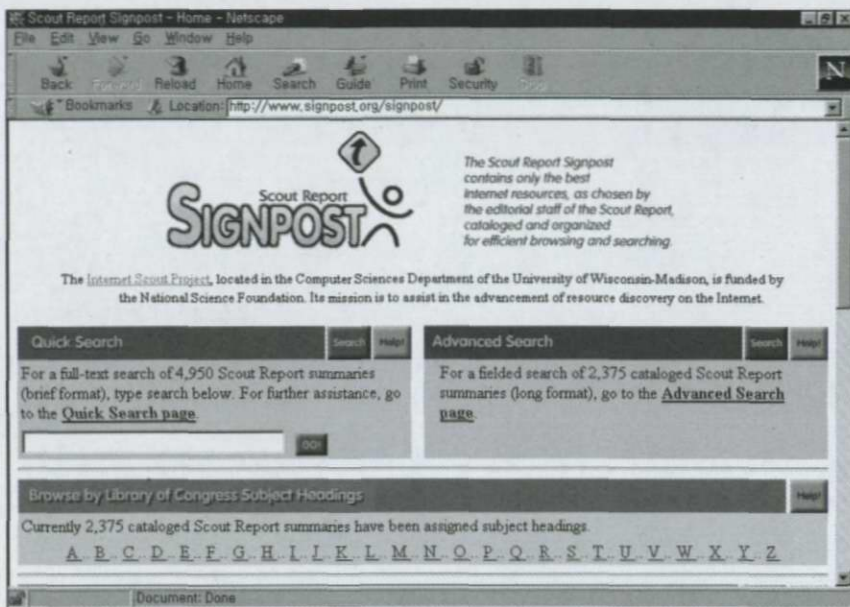
<예 2> HTML의 DC 표현 - LINK 태그

<pre><!ENTITY % a.global ' type CDATA #IMPLIED scheme CDATA "uncontrolled" > <!ELEMENT title - O (#PCDATA) > <!ATTLIST title %a.global > <!ELEMENT author - O (#PCDATA) > <!ATTLIST author %a.global > <!ELEMENT otherAgent - O (#PCDATA) > <!ATTLIST otherAgent %a.global > : <!ELEMENT coverage - O (#PCDATA) > <!ATTLIST coverage %a.global > <!ELEMENT metadata - O (#PCDATA) > <!ATTLIST metadata %a.global</pre>	<pre>X!ELEMENT dublinCore - O (title author otherAgent publisher date subject objectType form identifier relation source language coverage metadata)* > <!ATTLIST dublinCore version CDATA #IMPLIED ></pre>
---	--

<예 3> DC를 위한 SGML DTD

```
<!DOCTYPE dublinCore PUBLIC "-//OCLC//DTD Dublin core v.1//EN">
<dublinCore>
<title>On the Pulse of Morning</title>
<author>Maya Angelou</author>
<publisher>University of Virginia Electronic Text Center</publisher>
<otherAgent name="transcriber">University of Virginia Electronic Text Center</otherAgent>
<date name="created" scheme="ISO">1993-01-23</date>
<objectType>poem</objectType>
<form>1 ASCII file</form>
<form scheme="IMT">text/ASCII</form>
<source>Newspaper stories and oral performance of text at the Presidential inauguration of Bill Clinton</source>
<language name="ISO 639">en</language>
</dublinCore>
```

<예 4> DC의 SGML 마크업 예



(그림 2) Scout Signpost 초기화면

한정어를 정의하였고 문제시되는 데이터 요소에 대해서는 실무 작업반(working group)을 구성하여 계속적인 연구 작업을 수행하도록 하였다.

현재 15개 각 데이터 요소에 대해 제안되어 있는 한정어 가운데 Sub-element와 Scheme의 내용을 정리하면 <표 3>과 같고, 이에 대한 논의는 계속되고 있다.

5. DC 구문 기술

DC-2를 통해 DC의 구체적인 구문을 위한 연구가 진행되었다. DC의 구체적인 구문으로는 HTML 표현과 SGML DTD 개발을 통한 SGML 표현을 들 수 있다. 현재 HTML을 이용한 DC 표현은 HTML 버전 4.0까지 확장되었고, SGML 표현은 SGML의 하부집합인 XML을 이용하는 RDF(Resource Description Framework)에서 DC의 표현을 포괄하고 있다.

· HTML을 이용한 DC 표현

HTML은 메타데이터를 표현하기 위해 META와 LINK 태그를 사용하는데 메타데이터를 생성하기 위해서는 이들을 HTML 문서의 HEAD부분에 포함시켜야 한다.

META 태그의 경우, 이 태그가 갖는 속성으로 NAME(요소명)과 CONTENT(요소값)를 쓰게 되는데, 요소명은 'DC. 요소명'의 형식으로 표현한다. 한정어를 취하는 경우, 속성인 CONTENT 속에 Scheme과 Lang의 한정어를 넣어 표현할 수도 있고, 한정어를 태그의 속성으로 하여 나타낼 수도 있다.

LINK 태그는 HTML 문서내에 포함된 메타데이터를 식별하여 연결해 주는 것으로, 주로 URL이 사용된다. LINK 태그를 사용하면 DC로 표현된 다른 메타데이터와의 간접적인 연결이 가능하고, 이는 DC

가 아닌 다른 메타데이터 체계에서도 동일하게 사용될 수 있다.

· SGML을 이용한 DC 표현

DC-2에서 이루어진 DC의 구체 구문과 관련된 논의에서는 SGML 구문을 이용하여 DC를 사용할 수 있는 DTD의 개발이 제안되었다. DTD는 DC의 모든 메타데이터 요소의 속성 등을 정의하고 정의된 요소를 실제 SGML 문서에서 태그로 나타낸다. 다음 예시는 DC-2에서 제안한 SGML DTD의 일부와 실제 마크업 예를 나타낸 것으로, DC-2 당시의 13개 데이터 요소가 적용되었다.

6. DC 응용 사례

· BIBLINK (<http://hosted.ukoln.ac.uk/biblink/>)

BIBLINK 프로젝트는 유럽 연합과 여러 국가 도서관이 참여하는 프로젝트로서, 출판업자와 국가서지기관(NBA)간에 새롭게 출판되는 자료의 메타데이터 레코드를 상호 교환할 수 있도록 전자적 메타데이터 링크를 구축하는 것을 목표로 한다.

1996년 4월에 시작된 이 프로젝트는 3년간 계속되며, DC를 포함한 여러 가지 메타데이터에 대한 연구가 수행되어 표준이 되는 교환 포맷을 개발하고 이를 국가 도서관에서 사용하는 MARC로 전환할 수 있도록 추진하고 있다.

· Scout Report Signpost (<http://www.signpost.org/signpost/index.html>)
인터넷 Scout 프로젝트를 통해 개발된

Signpost는 주간 단위로 교육·과학 분야의 인터넷 자원을 선별하여 구축된 Scout Report를 탐색하고 검색하는 인터넷 서비스이다. 인터넷 자원을 기술하기 위해 DC를 이용하고 이를 기본으로 하여 자료의 목록, 분류를 위해 AACR2, LCSH, LCC를 이용하고 있다.

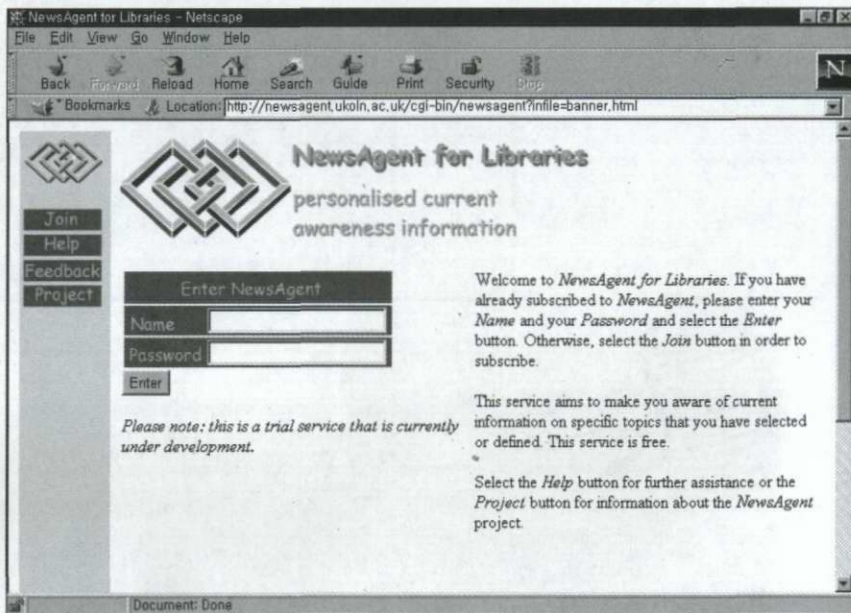
· NewsAgent for Libraries (<http://newsagent.ukoln.ac.uk/>)

NewsAgent for Libraries는 영국의 전자도서관 프로그램(eLib)의 전자 저널 부문의 프로젝트로서 도서관 및 정보관련 업무 실무자에게 전자 뉴스와 최신 정보 주지 서비스를 제공하기 위한 것이다.

현재 시범서비스가 이루어지고 있으며, Program, VINE, Library Technology, Ariadne, Journal of Librarianship and Information Science를 포함한 영국의 저명한 저널과 영국도서관협회, 영국국립도서관, UKOLN, LITC 등이 제공하는 뉴스와 브리핑 자료가 서비스에 포함되어 있다. NewsAgent에서 이용하는 자료의 메타데이터는 DC를 사용하여 기술하게 되는데, 현재 UKOLN은 이와 관련하여 서버가 자동으로 DC로 기술된 정보를 추출해내는 시스템을 개발하고 있다.

· EULER (<http://www.emis.de/projects/EULER/>)

EULER(European Libraries and Electronic Resources in Mathematical Sciences)는 수학분야에 관심있는 사용자가 관련 자료를 통합된 네트워크 환경에서 윈스톱으로 접근할 수 있도록 하는 프



〈그림 3〉 NewsAgent for Libraries 초기화면

로젝트로서 EU에서 수행하는 여러 가지 프로젝트 가운데 하나이다. 1998년 4월에 시작되어 2년반 동안 진행될 예정이다. DC를 이용하여 EULER가 제공하게 될 서지 데이터베이스, OPAC, 전자 저널, 회색문헌의 온라인 아카이브, 그리고 수학과 관련된 인터넷 자료의 색인 등을 통합할 메타데이터를 기술하고 EULER 엔진을 통해 탐색하도록 한다.

7. 발전 방향

1995년 DC-1이 개최되면서 DC 개발에 착수한 이래 메타데이터의 핵심적인 15개의 데이터 요소가 확정되었고, 또한 이들의 의미 확장과 Z39.50 프로토콜에서의 이용, 각종 메타데이터의 통합 방안에 이르기까지 많은 연구가 진행되어 왔다.

현재까지도 의미 확장을 위한 각 데이터 요소의 세부적인 사항들이 논의되고

있으며, Z39.50 프로토콜의 탐색 질의에서 사용되는 Bib-1 속성 집합과 DC의 각 데이터 요소와의 매핑 방안이 실무 작업반을 통해 연구되고 있다. 메타데이터 통합 방안으로는 DC-2에서 발표된 워릭 프레임 워크 이후에 RDF(Resource Description Framework) 등이 제안되고 있다.

이밖에 자료 기술의 표준으로서 비교적 상세하게 자료를 표현할 수 있는 메타데이터인 MARC, TEI, IAFI 템플릿과의 연계 혹은 매핑 문제도 끊임없이 연구되고 있다. 1998년 11월로 예정된 DC-6에서는 이렇게 다양한 DC의 응용 부문과 이와 밀접하게 관련된 다른 메타데이터와의 상호운용성을 높이는데 주력하게 될 것이다. ●