

Introduzione ai lavori del Gruppo di studio sui metadati descrittivi

Cristina Magliano

Il Gruppo di lavoro sui metadati descrittivi ha iniziato l'attività partendo dall'analisi del primo standard di metadati sviluppato da una Task Force di utilizzatori, cioè il Dublin Core Metadata Element Set (DCMES). Ricordiamo che DCMES stabilisce un vocabolario semantico per descrivere informazioni sulle caratteristiche "core" di un oggetto web e categorizzarlo ai fini di una ricerca semplificata da parte dell'utente. Fin dall'inizio però il consenso nella comunità del DC ritenne utile il suo utilizzo nell'ambito sia dell'oggetto digitale che di quello reale.

In questi ultimi tempi infatti l'estensione e lo sviluppo di alcuni metadati ha portato la loro applicabilità non solo alle risorse elettroniche o digitali ma anche alla descrizione di un qualsiasi tipo di risorsa quale ad esempio oggetti di un museo, di una biblioteca, di un archivio, etc. In alcune situazioni specifiche la risorsa può essere una collezione di oggetti, o una parte di essi: fra le tipologie degli elementi vi possono essere immagini, suoni, servizi, eventi, collezioni, testi, etc. Il metadato può descrivere il modello di un processo o di un evento., persone e ruoli, etc.

Si sono analizzate alcune realizzazioni già in atto, ad esempio il Mapping Dublin Core / Unimarc realizzato dall'ICCU e proposto per il progetto European Network in Europe di cui è partner. Il progetto, finanziato in parte dalla European Commission's Libraries Programme ONE-2, naturale prosecuzione del precedente progetto ONE, ha implementato un s/w in una architettura c/s basandosi per le procedure search and retrieval sul protocollo Z3950.

Sulla base di tabelle di comparazione (crosswalks) si è andata stabilendo nei diversi ambiti una equivalenza che tenga conto dei problemi inerenti la granularità e la semantica dei dati, mantenendo e non appiattendolo la specificità degli stessi. (es. titolo e type)

Sono stati inoltre analizzati i vari crosswalks già realizzati e pubblicati su Internet:

MetaForm Crosswalks, Crosscuts and Mappings - Microsoft Internet Explorer

Indietro Indirizzo <http://www2.sub.uni-goettingen.de/metaform/crosswalks.html> Vai

Collegamenti File Modifica Visualizza Preferiti Strumenti ?

METAFORM

CROSSWALKS, CROSSCUTS & MAPPINGS

Database containing Dublin Core manifestations and other metadata formats

Format	Project description	Version	M
<i>Dublin Core and its Dialects</i>			
Recordkeeping Metadata Standard for Commonwealth Agencies	Nationale Archives of Australia	Stand: 12.06.2000	M
BIBLINK	BIBLINK Project	Stand: 28.07.1998	M
BSZ	Dublin Core in der Interpretation des BSZ (Bibliothekservice-Zentrum Baden-Württemberg)	Version 1 Stand: 10.11.1999	M
CIC	Gesellschaft Deutscher Chemiker Keine Formatbeschreibung vorhanden		M
CORC	Cooperative Online Resource Catalog	Stand: 13.04.1999	M
	Deutsche Gesellschaft für Erziehungswissenschaft - Deutscher		

Internet

Metadata Standards Crosswalk

PRINT

CLOSE

*NOTE: To see the full list of categories for any one of the standards below, click on the name of the standard.
To print from your Web browser, use the landscape print format.*

CDWA	VRA 2.0 CoreCategories	VRA 3.0 CoreCategories	USMARC	Dublin Core	Object ID	FDA Guide	
OBJECT/WORK (core)							
Object/Work - Catalog Level						Document Classification- Catalogue Level (core) Document Classification-Group Type	
Object/Work-Type (core)	W1. Work Type	Type	655 Genre-Form	Type	Type of Object	Document Classification- Document Type (core) Purpose-Purpose (Broad) (core) Purpose-Purpose (Narrow)	object t
Object/Work- Components			300a Physical Description-Extent	Format.Extent		Document Classification-Extent	quantity

Programmi



MARC to Dublin Core Crosswalk - Microsoft Internet Explorer

Indirizzo <http://www.loc.gov/marc/marc2dc.html> Vai

Collegamenti >> File Modifica Visualizza Preferiti Strumenti ?

II. MARC to Dublin Core Crosswalk (Unqualified).

Conventions: "\$" is used to represent the control character subfield delimiter.

DC Element	MARC Fields	Implementation notes
Title	245	
Creator	100, 110, 111, 700, 710, 711	See Appendix 1 below; Contributor element not used.
	720	
Subject	600, 610, 611, 630, 650, 653	
Description	500-599, except 506, 530, 540, 546	
Contributor		See Appendix 1 below; Contributor element not used.
Publisher	260\$a\$b	
Date	260\$c	
Type	Leader06, Leader07	See Appendix 2 for Leader-Type rules
	655	
Format	856\$q	
Identifier	856\$u	
Source	786\$o\$t	
Language	008/35-37	
	546	

Operazione completata Internet

MARC to Dublin Core Crosswalk - Microsoft Internet Explorer

Indietro >> Indirizzo <http://www.loc.gov/marc/marc2dc.html> Vai

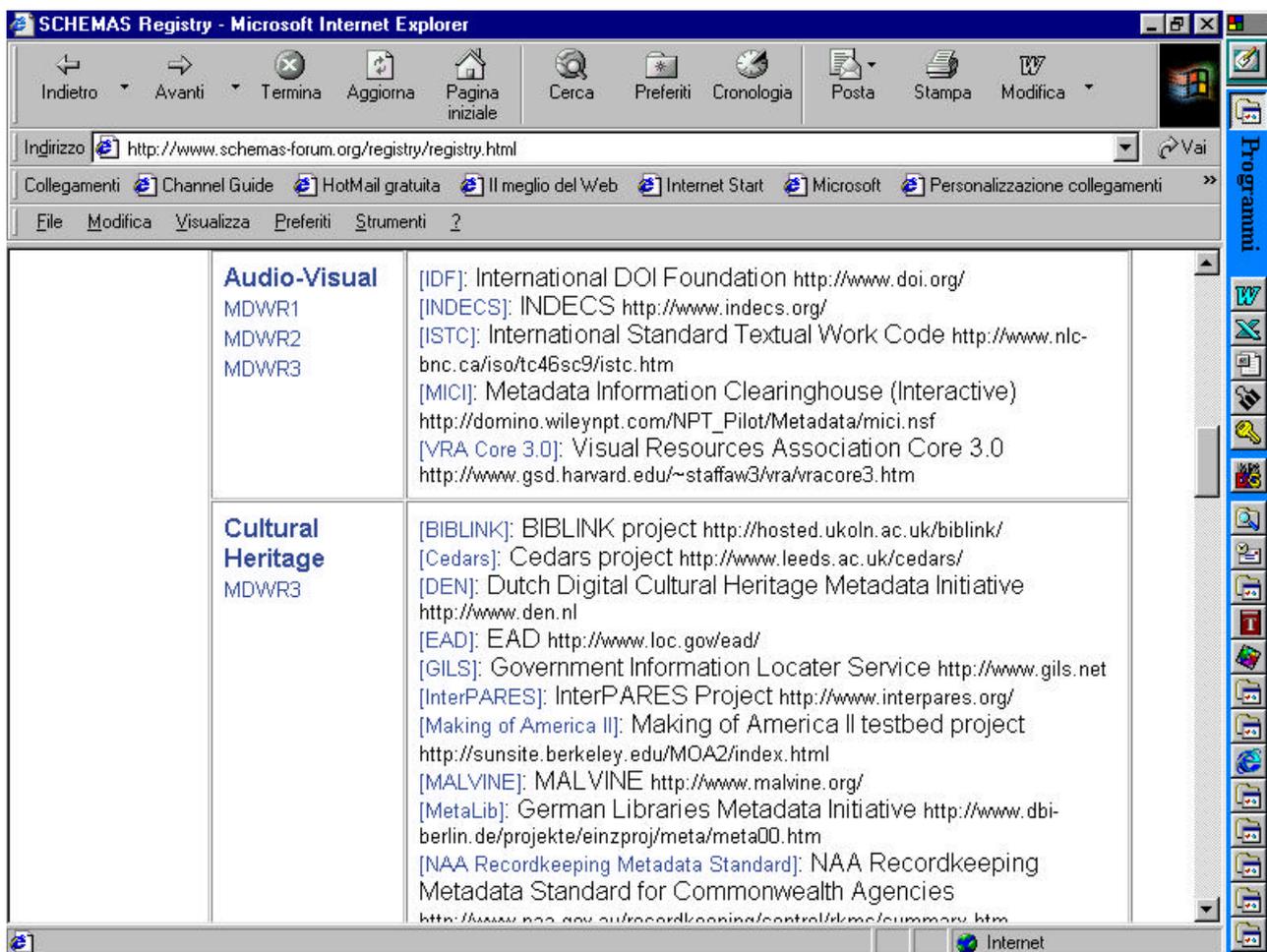
Collegamenti >> File Modifica Visualizza Preferiti Strumenti ?

III. MARC to Dublin Core Crosswalk (Qualified).

DC Element	DC Qualifier(s)	MARC Fields	Implementation notes
Title		245	
Title	Alternative	130, 210, 240, 242, 246, 730, 740	
Creator		100, 110, 111, 700, 710, 711 720	See Appendix 1 below.
Subject	LCSH	600, 610, 611, 630, 650	Second indicator=0
Subject	MeSH	600, 610, 611, 630, 650	Second indicator=2
Subject	LCC	050	
Subject	DDC	082	
Subject	UDC	080	
Description		500-599, except 505, 506, 520, 530, 540, 546	
Description	TableofContents	505	
Description	Abstract	520	First indicator=3
Contributor			See Appendix 1 below; Contributor element not used.
Publisher		260\$a\$b	
Date	Created	260\$a\$b	

Operazione completata Internet

Mapping della LC con il formato MARC 21, sia nella versione con qualificatori che senza.



Il processo di standardizzazione dei metadati (modelli, semantica e sintassi) è iniziato ma come si può vedere all'interno di domini specifici. Un modello generale RDF (Resource Description Framework) è stato proposto con una implementazione in linguaggio XML (eXtended Markup Language), ma non sembra avere ottenuto un consenso ed una accettazione diffusa, anche perché definizioni valide per tutti gli ambiti non sono facilmente applicabili.

Si è quindi stabilita una strategia di mappatura che rispettasse le specificità e tenesse anche conto degli standard descrittivi e dei profili esistenti, per applicare un modello il più possibile scalabile che interagisca fra sistemi informativi eterogenei. Il gruppo ha predisposto un modello per la raccolta dei dati descrittivi per individuare gli elementi ritenuti essenziali alla identificazione dell'oggetto digitale ai fini di una condivisione delle informazioni descrittive tra istituzioni che utilizzano formati catalografici differenti. Nella definizione di un set minimo di dati è sembrato, fin dove è stato possibile e applicabile, fare riferimento ad un modello logico e alle indicazioni fornite dallo studio dell'IFLA "Functional Requirements for Bibliographic Records", applicando la lista di attributi e relazioni nella descrizione delle risorse digitali, secondo le quattro funzioni fondamentali dei record bibliografici:

trovare, identificare, selezionare, ottenere

Fin dall'inizio è emersa la complessità della costituzione di un formato di metadati per una vasta comunità di utenti. D'altra parte l'utilizzo di informazioni dettagliate come quelle di un formato bibliografico sembravano essere troppo lontane e non appropriate per essere

selezionate per una ricerca di oggetti digitali ad un livello di cross-domain e di cooperazione. Inoltre gli interessi degli utenti/e o produttori di una risorsa digitale, poco stabile e tecnologicamente aggiornabile, "necessita di informazioni specifiche, spesso difficili da recuperare, che può solo basarsi su un lavoro di integrazione non solo fra diversi attori, ma anche fra servizi, strumenti tecnologici e meccanismi amministrativi che finora, nel trattamento del materiale tradizionale, erano separati fra loro e comunque meno complessi da gestire perché già consolidati" (vedi Studio di fattibilità per la realizzazione della Biblioteca Digitale).

Illustrazione dello schema

Elemento Title e Type

In questo primo mapping (<http://www.iccu.sbn.it/mapmeta1.doc>) si sono confrontati i dati bibliografici con quelli delle opere d'arte, disegni, numismatica, stampe e matrici di incisioni, folklore, reperti archeologici, edifici e manufatti architettonici, risorse elettroniche locali e remote, materiale d'archivio, documentazione ufficiale (resoconti e atti parlamentari). Si è partiti dagli elementi del set DC per effettuare un primo confronto e una prima verifica in quanto primo standard consolidato per una analisi semantica ed una descrizione sulle caratteristiche "core" di un oggetto web ed anche per una categorizzazione ai fini della sua reperibilità.

Prima indicazione data dal gruppo è stata quella di attenersi a standard consolidati nazionali e internazionali ed eventualmente adattarli agli scopi e alle esigenze specifiche

Partendo dalla struttura dei dati e dal modo in cui questi sono utilizzati nei vari domini (Arti, Archivi, Biblioteche, Enti governativi/Atti parlamentari), lo schema ha previsto per ciascun elemento, sia gli standard/profilo di MD utilizzati, sia gli standard catalografici specifici con le relative occorrenze. Il primo approccio è stato quello di confrontare gli elementi descrittivi, dal punto di vista semantico ed anche sintattico fra le diverse applicazioni.

Analizzando i vari standard catalografici, come si può vedere dall'esempio proposto, una diversa accezione riguarda proprio l'elemento di ricerca più comune e diffuso, almeno nell'ambito delle biblioteche: Title.

Partendo dalle relative definizioni nei diversi ambiti si evidenziano le diverse accezioni dell'elemento Title, dove, in ambito storico artistico può riferirsi sia a nomi storici o tradizionali dell'oggetto (per reperto archeologico e ed opere ed oggetti d'arte), sia a denominazioni del bene da fonti ufficiali (architettura), sia a dati relativi al soggetto secondo le definizioni iconografiche tradizionali.

In ambito archivistico corrisponde al titolo dell'unità come sottoelemento dell'identificativo della descrizione documentaria.

Nelle biblioteche corrisponde sia al titolo proprio che al complemento del titolo, includendo altri titoli di accesso quali le forme varianti, il titolo uniforme ed il titolo chiave.

L'elemento Type presenta varie accezioni: in ambito bibliografico e storico artistico infatti può descrivere sia la designazione generica del materiale, sia la designazione specifica. In ambito archivistico indica sostanzialmente il livello gerarchico dell'unità documentaria che si sta descrivendo.

Nella stesura finale lo schema conterrà le estensioni di tutti gli elementi nei diversi ambiti, la scelta di utilizzarli dipenderà dalla tipologia dell'utenza, dal servizio che si intende fornire

e dalla disponibilità di personale esperto in grado di compiere una analisi adeguata degli elementi e di definirli secondo criteri di standardizzazione e conoscibilità.

Inoltre saranno individuati, soprattutto per le modalità di ricerca che necessitano di standardizzazione, liste normalizzate, quali ad es. i codici di ruolo delle responsabilità da prevedere, partendo dal confronto con i relator code di "Unimarc - Appendix C"; norme UNI/ISO, UNI/ISO 5963-189 per l'indicizzazione, ISO 214-1976 per l'abstract, liste controllate TGN geografica e W3CDTF, temporale per l'elemento coverage, etc. Infatti il ricorso a tali strumenti come authority files, schemi di classificazione, sistemi di soggettazione, è un elemento qualificante di qualsiasi progetto di digitalizzazione e sono strettamente necessari ai fini di garantire l'interoperabilità.

Sulla base delle diverse semantiche si sta approntando inoltre un glossario dei termini usati con la definizione e la relativa indicazione dei diversi standard catalografici di riferimento.

Il Gruppo di studio ha individuato ulteriori settori che dovranno essere affrontati per lo sviluppo di uno standard di metadati:

- L'interoperabilità linguistica (Thesauri) e tecnica (formato ed estensione dei file)
- La disponibilità di dati normalizzati ai fini dell'utilizzo da parte dei produttori e degli utenti finali
- L'individuazione degli elementi univoci di identificazione e di accesso (n. standard, segnatrice, URN etc.).