

InformaticaUmanistica

Lezione 8

Introduzione ai Metadati, modello MARC e Dublin Core

Pasquale Savino

ISTI - CNR



UNIVERSITÀ DI PISA

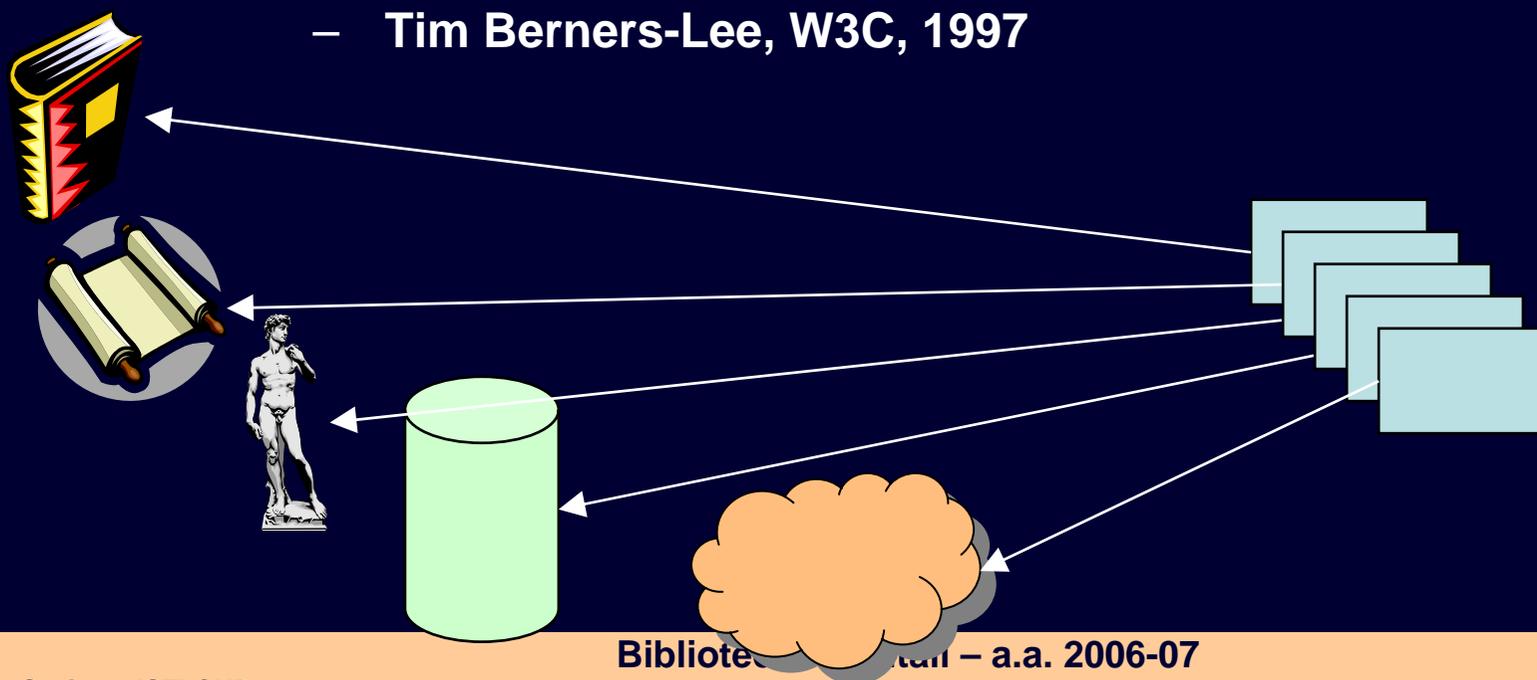
Sommario

- ◆ **Definizione**
- ◆ **I vari tipi di metadati**
- ◆ **Modelli di metadati**
 - MARC21
 - Dublin Core
 - MPEG-7
 - ECHO

Metadati - Definizione

Cosa sono i metadati?

- Alcune definizioni semplici...
- 'Structured data about data'.
 - **Dublin Core Metadata Initiative FAQ, 2003**
- Machine-understandable information about Web resources or other things.
 - **Tim Berners-Lee, W3C, 1997**



Una visione “funzionale” dei metadati

- Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics.
A user might be a program or a person.
 - **Lorcan Dempsey & Rachel Heery, 1998**
- Structured data about resources that can be used to help support a wide range of operations
 - **Michael Day, 2001**

A quali risorse possono essere associati dei metadati?

- I metadati possono essere associati a qualunque cosa – digitale, fisica, risorsa “astratta”
 - HTML documents
 - digital images
 - databases
 - books
 - museum objects
 - archival records
 - metadata records
 - Web sites
 - collections
 - services
 - physical places
 - people
 - institutions
 - abstract “works”
 - concepts
 - events

Che informazione contengono i metadati?

- I record di metadati includono
 - ➔ Record bibliografici appartenenti a cataloghi o ottenuti da servizi di indicizzazione e creazione di abstract
 - ➔ Descrizione del materiale archiviato
 - ➔ Record relativi ad oggetti contenuti nella documentazione del museo
 - ➔ Descrizione di oggetti digitali (documenti, immagini, video software, ...)
 - ➔ Descrizione di collezioni di oggetti digitali
 - ➔ Descrizione di servizi in rete
 - ➔ Descrizione di record di metadati

Quali operazioni possono essere effettuate sui metadati?

- Operazioni da parte degli utenti o da parte di tool software
- I metadati possono essere usati per supportare diverse funzioni, quali ad esempio
 - ➔ resource disclosure & discovery
 - ➔ resource management, (incluso preservation)
 - ➔ intellectual property rights management
 - ➔ Autenticazione ed autorizzazione
 - ➔ Personalizzazione e localizzazione dei servizi
- Le diverse funzioni richiedono diversi tipi e classi di metadati
 - ➔ Non è possibile trovare una soluzione unica per tutti gli usi
 - ➔ È necessario specificare i requisiti funzionali per definire le caratteristiche dei metadati da utilizzare

Elementi ed insiemi di elementi dei Metadati

- I Metadati descrivono gli attributi e le proprietà di una risorsa
- Ogni attributo o proprietà è descritto da un **metadata element**
 - Può essere identificato e definito in modo preciso
 - Ogni metadata element può essere rappresentato in diverse forme
- Un **metadata element set** è
 - Insieme coerente di elementi
 - Creato per uno scopo specifico, come una singola unità
- Schema
 - Rappresentazione strutturata di un element set

Autore = J Smith
Data = 2001-11-05
Titolo = Report

Resource1

I tipi di metadati

Tipi di metadati

- ◆ **Metadati descrittivi o per “resource discovery”**
 - Utilizzati per descrivere le risorse della Biblioteca Digitale
 - Utilizzati per facilitare la ricerca di documenti nella BD
 - Cataloghi, indici, ecc.
- ◆ **Metadati strutturali**
 - Descrivono la struttura degli oggetti della BD e le relazioni che esistono tra oggetti (o loro parti) ed altri oggetti
- ◆ **Metadati amministrativi**
 - Utilizzati per la gestione delle collezioni di documenti e per il controllo dell'accesso

Metadati per resource discovery

- Gli utenti desiderano
 1. Individuare le risorse in accordo a determinati criteri
 - Ad es. *Trova tutti i documenti scritti da ... che parlano di ...*
 2. (oppure) identificare una specifica risorsa
 - Essere in grado di distinguere tra diverse risorse simili tra loro
 3. Selezionare
 - **Scegliere una risorsa che sia appropriata ai bisogni**
 4. Localizzare la risorsa
 5. *Ottenere/accedere alla risorsa*
 6. *Utilizzare la risorsa*
 - ***open, read, display, run, play, copy, unpack/repack***
 - ***Interpretare il contenuto***
- I metadati per resource discovery sono utilizzati principalmente per le operazioni 1-4

Metadati per resource discovery

Continuum di complessità/funzionalità

Indici full-text, feature per immagini	Generato con tool software in modo automatico. Secondo alcune definizioni potrebbero non essere considerati metadati	discovery per contenuto, localizzazione
Forme semanticamente semplici (e.g. Dublin Core)	Permettono la descrizione di un gran numero di risorse In alcuni casi possono essere generati in modo automatico, in altri sono prodotti manualmente	discovery, identificazione, selezione, localizzazione
Forme complesse più ricche (e.g. MARC, ECHO, etc)	Relativo a tipi di risorse specifiche Spesso associato con domini/comunità specifici La creazione può richiedere un significativo intervento umano	discovery, identificazione, selezione, localizzazione, accesso, uso

Associazione di una risorsa ai metadati [1/3]

Creator = J Smith
Date = 2001-11-05
Title = Report

Resource1

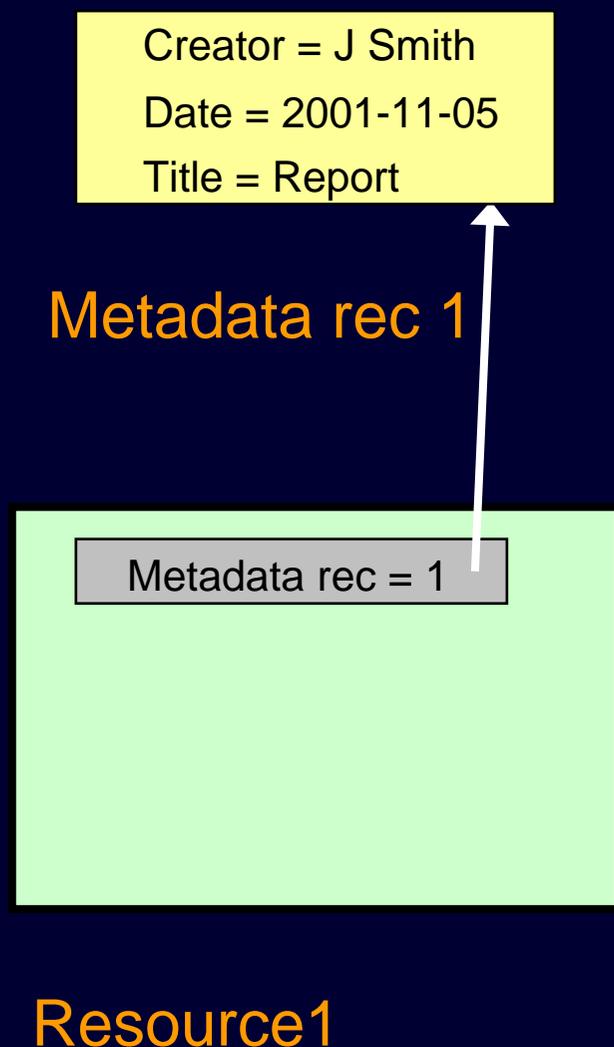
I metadati possono essere contenuti all'interno della risorsa

e.g. elementi `meta` in doc HTML;

Proprietà del documento nei word processor

- Le risorse possono permettere che i metadati siano contenuti al loro interno (embedding)?
- Il creatore dei metadati può avere i diritti di **scrittura** sulla risorsa?
- Può un servizio **estrarre** i metadati che sono contenuti nella risorsa?
- Possono esistere metadati per **aggregati di risorse**? Come vengono gestiti?

Associazione di una risorsa ai metadati [2/3]



*Il record dei metadati è un oggetto separato
L'identificatore del record e' contenuto
nella risorsa*

e.g. elementi `link` elements in doc HTML I
record di metadati possono essere remoti
rispetto alla risorsa

- Le risorse devono poter contenere i link
- Il creatore dei metadati può avere i diritti di **scrittura** sulla risorsa?
- Devono esistere dei meccanismi per seguire i link
- Cosa succede se una risorsa viene cancellata?
- Possono esistere metadati per **aggregati di risorse**? Come vengono gestiti?

Associazione di una risorsa ai metadati [3/3]

Doc = 1
Creator = J Smith
Date = 2001-11-05
Title = Report

Metadata rec 1



Resource 1

***Il record dei metadati è un oggetto separato
L'identificatore della risorsa e' contenuto nel
record***

Il record dei metadati può essere remoto rispetto alla risorsa

Non richiede che nessuna informazione sui metadati sia contenuta nella risorsa Non richiede che il creatore dei metadati possa modificare la risorsa

Il record dei metadati può essere creato indipendentemente dalla risorsa – possibilità di record multipli

I servizi utilizzano i metadati indipendentemente dalla risorsa

Il record dei metadati può continuare ad esistere anche se la risorsa viene cancellata

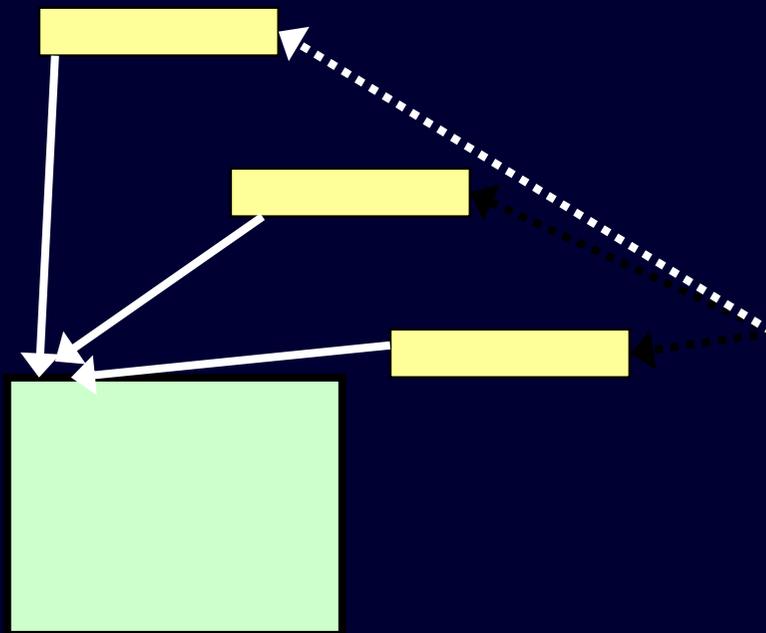
Il record dei metadati può descrivere qualunque oggetto che abbia un identificatore

Gestione dei metadati

Il record dei metadati viene usato separatamente dalla risorsa che viene descritta

I metadati sono essi stessi una risorsa

Il contenuto dei metadati può essere memorizzato nella forme più appropriate per il servizio: in un database, in forms, ecc.



Doc	Creator	Date	Title
1	J Smith	2001-11-05	Report

MARC

MARC – Introduzione

- ◆ **Sviluppato alla fine degli anni 60 alla Library of Congress (USA) per promuovere la condivisione di cataloghi tra biblioteche**
- ◆ **È diventato un formato quasi standard**
- ◆ **Utilizza un sistema di numeri, lettere e simboli all'interno del record per individuare i diversi tipi di informazione**
- ◆ **Fornisce**
 - Descrizione dell'oggetto (titolo, edizione, pubblicazione, ecc.)
 - Soggetto
 - Numero di catalogo
- ◆ **Molti formati simili a MARC sono stati sviluppati da organismi bibliotecari nazionali (USMARC, UKMARC, CANMARC, AUSMARC, DANMARC, ANNAMARC, INTERMARC, ...)**
- ◆ **USMARC e CANMARC sono stati pubblicati in una edizione unificata nel 1999 con il nome di MARC 21**
- ◆ **UNIMARC (Universal MARC) come formato standard e come formato di scambio tra i diversi formati**
 - Ad es. USMARC → UNIMARC → INTERMARC
 - Problemi di degrado dell'informazione nei vari passaggi

Descrizione generale

- ◆ L'informazione contenuta in ogni record MARC deriva dal formato delle schede di catalogazione e dipende dal particolare oggetto che si vuole descrivere
- ◆ Deve essere organizzata in modo che possa essere interpretata da un programma, quindi le diverse informazioni devono essere codificate precisamente.
- ◆ È necessario permettere una certa flessibilità nel formato, per poter descrivere oggetti diversi
 - Ad esempio, il titolo di un'opera avrà una lunghezza variabile
 - Vi sono degli oggetti che sono parte di una serie, mentre altri non lo sono
 - Ecc.
- ◆ Ogni record è comunque composto
 - “tabella dei contenuti”
 - Descrizione dell'oggetto
 - Un insieme di **soggetti** selezionati da una lista predefinita
 - Un **numero di collocazione** che indica dove si trova l'oggetto

Record della Library of Congress

LC Control Number:68093427

Type of Material:Book (Print, Microform, Electronic, etc.)

Personal Name:[Manzoni, Alessandro, 1785-1873.](#)

Main Title:I promessi sposi.

Edition Information:[2. ed.]

Published/Created:Firenze, La nuova Italia, 1967.

Related Names:[Russo, Luigi, 1892-1961. ed.](#)

Description:xxii, 716 p. 21 cm.

LC Classification:PQ4713 .P3 1967

Other System No.:(OCoLC)23497169

CALL NUMBER:

[PQ4713 .P3 1967](#)

Copy 1-

Request in:Jefferson or Adams Bldg General or Area Studies Reading Rms-

Status:Not Charged

Record della Library of Congress MARC

LC Control Number:68093427
000 00678nam 22002291 450
001 8495557
005 19980421195046.0
008 890330s1967 it 000 0 ita
035 __ |9 (DLC) 68093427
906__ |a 7 |b cbc |c oclcrpl |d u |e ncip |f 19 |g y-gencatlg
010 __ |a 68093427
035 __ |a (OCoLC)23497169
040 __ |a DLC |c MiU |d DLC
050 00 |a PQ4713 |b .P3 1967
100 1_ |a Manzoni, Alessandro, |d 1785-1873.
245 12 |a I promessi sposi.
250 __ |a [2. ed.]
260 __ |a Firenze, |b La nuova Italia, |c 1967.
300 __ |a xxii, 716 p. |c 21 cm.
700 1_ |a Russo, Luigi, |d 1892-1961. |e ed.
985 __ |e OCLC REPLACEMENT
991 __ |b c-GenColl |h PQ4713 |i .P3 1967 |t Copy 1 |w OCLCREP

Struttura dei record

Ogni record MARC è composto di una intestazione, un insieme di campi di controllo ed un insieme di campi con i dati.

`<leader> <directory> FT <control-number-field> FT [<control-field>]+ FT [<data-field>]+ FT RT`

- ◆ `<leader>` - campo a lunghezza fissa che contiene informazione generale sul record
- ◆ `<directory>` - indice alla posizione dei campi di controllo e dati (`<control-field>` e `<data-field>`) all'interno del record
- ◆ `<control-number-field>` - carattere di controllo specifico dell'organizzazione che ha generato il record MARC
- ◆ `<control-field>` - contiene informazione di controllo specifica del record
- ◆ `<data-field>` - contiene i valori di metadati
- ◆ FT - indica la terminazione di un campo
- ◆ RT - indica la terminazione del record

Intestazione

- ◆ **L'intestazione contiene informazione che normalmente viene utilizzata dal sistema**
 - Leader
 - 24 caratteri utilizzati dal sistema
 - Contiene informazioni quali la lunghezza del record, il tipo, ecc.
 - Directory
 - Specifica quali tag ci sono nel record e dove si trovano
 - Il campo 008
 - Contiene informazione importante ma in forma sintetica
 - Ad es. può contenere un codice per il paese di pubblicazione, un codice per indicare la lingua del testo, ecc.

Il campo <leader>

Composto di 24 byte di caratteri ASCII

0 – 4	record-length	Lunghezza del record
5	record-status	Stato del record (ad es. nuovo, modificato, ecc.)
6	type-of-record	Caratteristiche del record
7 – 8	implementation-defined	Riservato a specifiche implementazioni
9	character-coding-schema	Schema di codifica dei caratteri utilizzato
10	indicator-count	Numero di indicatori presenti in ogni campo dati
11	subfield-code-length	Lunghezza del delimitatore + identificatore dei dati
12 – 16	base-address-of-data	Posizione del primo carattere del campo variabile del record
17 – 19	implementation-defined	Riservato a specifiche implementazioni
20 – 23	entry-map	Specificano la struttura del campo directory

Il campo <directory>

- ◆ fornisce informazioni sui campi variabili del record. Insieme di elementi, di lunghezza fissa, uno per ogni campo variabile del record.

0 - 2	Tag	Tre caratteri numerici che identificano il campo
3 - 6	Length-of-field	Lunghezza del campo
7 - 11	Starting-character-position	Fornisce la posizione di inizio del campo

Struttura del record

- ◆ **<control-number-field> contiene un carattere di controllo utilizzato dall'organizzazione che ha inserito il record.**
- ◆ **I record MARC possono contenere uno o più campi <control-field>. Questi campi iniziano con un tag di tre cifre, avente le prime due poste a zero (quindi hanno la forma 00X).**
- ◆ **Il campo <data-field> contiene i dati del record MARC.**

Il campo <data-field>

- ◆ Ogni campo <data-field> inizia con un <field-tag> - un codice di 3 cifre di cui la prima è diversa da zero.
- ◆ Ad ogni <field-tag> corrisponde uno specifico tipo di informazione, quale ad esempio *autore*, *titolo*, *edizione*, ecc.

<field-tag> [<indicator-1>] [<indicator-2>]

[<delimiter> <subfield-tag>,<subfield-value>]+

- ◆ <indicator-1> ed <indicator-2> sono campi opzionali che specificano come deve essere interpretato il valore associato al <field-tag>
- ◆ Vi possono essere uno o più sotto-campi, composti da un delimitatore (<delimiter>), da un identificatore del sottocampo (<subfield-tag>) e dal valore corrispondente (<subfield-value>). Il <subfield-tag> è costituito da una lettera minuscola.

Esempi di field-tag

- ◆ **010 tag** marks the **Library of Congress Control Number (LCCN)**
- ◆ **020 tag** marks the **International Standard Book Number (ISBN)**
- ◆ **100 tag** marks a **personal name main entry (author)**
- ◆ **245 tag** marks the **title information**
- ◆ (which includes the title, other title information, and the statement of responsibility)
- ◆ **250 tag** marks the **edition**
- ◆ **260 tag** marks the **publication information**
- ◆ **300 tag** marks the **physical description** (often referred to as the "collation" when describing books)
- ◆ **440 tag** marks the **series statement/added entry**
- ◆ **520 tag** marks the **annotation or summary note**
- ◆ **650 tag** marks a **topical subject heading**
- ◆ **700 tag** marks a **personal name added entry** (joint author, editor, or illustrator)

Struttura di MARC

◆ Indicatori

- Ogni campo (con alcune eccezioni) può essere seguito da due caratteri usati come indicatori.
- Gli indicatori specificano come deve essere interpretato il valore associato al campo
- Gli indicatori possono assumere valori da 0 a 9 (# se l'indicatore è vuoto)
- Ad esempio

245 14 \$a The emperor's new clothes /
 \$c adapted from Hans Christian Andersen
 and illustrated by Janet Stevens.

- 245 è il campo per Titolo
- Il primo indicatore con valore 1 specifica che nel catalogo esiste una diversa scheda per il titolo
- Il secondo indicatore specifica il numero di caratteri all'inizio del campo che devono essere trascurati (nell'esempio sono 4)

Struttura di MARC

◆ Sottocampi

- Diversi campi sono costituiti da diverse parti di informazione. Ogni parte è specificata da un sottocampo, con uno specifico codice
- Ogni sottocampo è preceduto da un delimitatore
- Ad esempio

300 ## \$a 675 p. : \$b ill. ; \$c 24 cm.

300 è il campo “descrizione fisica di un libro”

Il delimitatore del sottocampo è \$

Vi sono 3 sottocampi

Numero di pagine (\$a)

Informazioni sulle illustrazioni (\$b)

Dimensioni (\$c)

Regole generali per la codifica dei field-tag

- ◆ **La divisione dei tag di MARC21 è la seguente**
 - **0XX** Informazione di controllo, numeri, codici
 - **1XX** Parte principale
 - **2XX** Titoli, edizione, statement di responsabilità, informazione sulla pubblicazione
 - **3XX** Descrizione fisica, etc.
 - **4XX** Serie
 - **5XX** Note
 - **6XX** Soggetto
 - **7XX** Altre entries diverse dal soggetto o dalla serie
 - **8XX** Altro
 - **9XX** Lasciato per uso locale

Punti d'accesso

- ◆ I punti d'accesso costituiscono le parti più importanti del record
- ◆ Vengono utilizzati per effettuare le ricerche dei record
 - 1XX fields (main entries)
 - 4XX fields (series statements)
 - 6XX fields (subject headings)
 - 7XX fields (added entries other than subject or series)
 - 8XX fields (series added entries)
- ◆ Questi campi devono rispettare dei formati prestabiliti (Authority control)

Authority control

- ◆ **Permette la scelta di forme codificate ed uguali per tutti. Un catalogatore sceglie i soggetti ed i nomi da una lista utilizzata da tutti**
 - Ad esempio, se in una intestazione parliamo alle volte di “Getty Museum” ed altre di “J. Paul Getty Museum“, avremo problemi in fase di ricerca
 - Per i nomi si utilizzano il Library of Congress Name Authority file
 - Per gli argomenti ed i nomi geografici si usano il *Library of Congress Subject Headings* o il *Sears List of Subject Headings*.

Regole generali per la codifica dei field-tag

- ◆ Ogni tipo di tag (da 0XX a 9XX) può specificare valori che appartengono a diverse tipologie di dati (Nomi di persona, Nomi di aziende, Nomi geografici, ecc.)
- ◆ I campi che richiedono un “authority control” avranno tag specifici per valori di tipologie specifiche
 - X00 Nomi di persona
 - X10 Nomi di aziende
 - X11 Nomi di meeting
 - X30 Titoli
 - X40 Titoli bibliografici
 - X50 Termini relativi ad argomenti
 - X51 Nomi geografici
- ◆ Ad es., se il soggetto di un libro (6XX) è una persona, il tag sarà 600, se è un'azienda il tag sarà 610

```

01041cam 2200265 a 4500001002000000000300040002000
50017000240080041000410100024000820200025001060200
04400131040001800175050002400193082001800217100003
20023524500870026724600360035425000120039026000370
04023000029004395000042004685200220005106500033007
30650001200763^###89048230#/AC/r91^DLC^19911106082
810.9^891101s1990####maua###j#####000#0teng##^##$
a###89048230#/AC/r91^###$a0316107514 :$c$12.95^##$a
0316107506 (pbk.) :$c$5.95 ($6.95 Can.)^##$aDLCScD
LC$dDLC^00$aGV943.25$b.B74 1990^00$a796.334/2$220^
10$aBrenner, Richard J.,$d1941-^10$aMake the team.
$bSoccer :$ba heads up guide to super soccer! /$cR
ichard J. Brenner.^30$aHeads up guide to super soc
cer.^##$alst ed.^##$aBoston :$bLittle, Brown,$ccl19
90.^##$a127 p. :$bill. ;$c19 cm.^##$a"A Sports ill
http://www.loc.gov/marc/umb/um11to12.html (1 di 14) [14/01/2004
17.11.32]

```

Leader

Tag directory

Field 001

Lunghezza tag

Inizio tag

Terminatore

```

Understanding MARC Bibliographic: Parts 11 to 12
ustrated for kids book."^##$aInstructions for impr
oving soccer skills. Discusses dribbling, heading,
playmaking, defense, conditioning, mental attitud
e, how to handle problems with coaches, parents, a
nd other players, and the history of soccer.^#0$aS
occer$vJuvenile literature.^#1$aSoccer.^

```

Esempio di record MARC

```

Leader          01041cam 2200265 a 4500
Control No.     001 ###89048230
Control No. ID 003 DLC
DTLT           005 19911106082810.9
Fixed Data     008 891101s1990 maua j 001 0 eng
LCCN           010 ## $a ###89048230
ISBN           020 ## $a 0316107514 :
                $c $12.95
ISBN           020 ## $a 0316107506 (pbk.) :
                $c $5.95 ($6.95 Can.)
Cat. Source    040 ## $a DLC
                $c DLC
                $d DLC
LC Call No.    050 00 $a GV943.25
                $b .B74 1990
Dewey No.     082 00 $a 796.334/2
                $2 20
ME:Pers Name   100 1# $a Brenner, Richard J.,
                $d 1941-
Title         245 10 $a Make the team.
                $p Soccer :
                $b a heads up guide to super soccer! /
                $c Richard J. Brenner.

```

Visualizzazione del record

TITLE : Make the team. Soccer : a heads up guide to super soccer! / Richard J.Brenner.

ADDED TITLE : Heads up guide to super soccer

AUTHOR : Brenner, Richard J., 1941-

PUBLISHED : 1st ed. Boston : Little, Brown, c1990.

MATERIAL : 127 p. : ill. ; 19 cm.

NOTE : "A Sports illustrated for kids book."

NOTE : Instructions for improving soccer skills. Discusses dribbling, heading, playmaking, defense, conditioning, mental attitude, how to handle problems with coaches, parents, and other players, and the history of soccer.

SUBJECT : Soccer--Juvenile literature.

Soccer.

Copies Available : GV943.25 .B74 1990

**Visualizzazione
del record**

Riferimenti

- ◆ **Understanding MARC**, Betty Furrie in conjunction with the Data Base Development Department of The Follett Software Company, Published by the Cataloging Distribution Service, Library of Congress, in collaboration with The Follett Software Company, <http://www.loc.gov/marc/umb/>
- ◆ *Anglo-American Cataloguing Rules*. 2nd ed., 2002 revision. Chicago: American Library Association, 2002.
- ◆ Dewey, Melvil. *Dewey Decimal Classification and Relative Index*. 22nd. 4 vols. Albany, N.Y.: Forest Press, 2003.

Dublin Core

Dublin Core: background

- ◆ Necessità di dare una rappresentazione formalizzata alle risorse di informazione accessibili in rete via WEB
- ◆ Varietà delle risorse: testi, suoni, immagini tridimensionali, video, fotografie, film ...
- ◆ Necessità di una rappresentazione capace di comprendere i vari tipi di risorse

Dublin Core: background

- ◆ Esigenze accolte da un gruppo di bibliotecari, e di informatici....
- ◆ “Finding relevant information on the World Wide Web has become increasingly problematic due to the explosive growth of networked resources. Current Web indexing evolved rapidly to fill the demand for resource discovery tools, but that indexing, while useful, is a poor substitute for richer varieties of resource description.”
- ◆ Dublin Core Metadata Initiative <http://www.ietf.org/rfc/rfc2413.txt>
- ◆ **Dublin Core Metadata Set. Un insieme di elementi descrittivi capaci di rappresentare qualsiasi risorsa di informazione accessibile in rete**

Dublin Core: background

- ◆ **Il *D.C. metadata set* e' il formato di metadati piu' diffuso oggi, risultato di un progetto iniziato nel 1995 [Dublin Core Metadata Workshop Series] con l'obiettivo di affrontare il problema della ricerca di risorse di informazione disponibili sulla rete**
 - documenti testuali, programmi, video....ma anche basi di dati, archivi, etc..
- ◆ **La sua caratteristica - frutto di una scelta ben precisa - è la generalità e la semplicità. Esso cioè è stato pensato come un insieme minimo di elementi descrittivi capaci di rappresentare qualsiasi risorsa di informazione -pagina web, un audio, un video etc. - che abbia un indirizzo WEB**

Caratteristiche di Dublin Core

- ◆ *Si applicano a tutti i tipi di risorse di informazione*
- ◆ *Sono facili da creare, anche da parte di personale non specializzato (a differenza di quanto succede per i record catalografici di tipo MARC)*
- ◆ *Sono di uso generale per la “scoperta” delle risorse di informazione (a differenza delle descrizioni catalografiche che in genere seguono normative specifiche del settore di applicazione: biblioteche, musei, videoteche...)*
- ◆ *Il modello Dublin Core è “piatto”, cioè non permette di descrivere metadati strutturati, ed ha un numero di elementi limitato (15)*

Linee guida nella definizione di Dublin Core

◆ Semplicità

- Anche i non esperti devono poter creare e gestire i metadati
- Insieme di metadati sufficiente per una ricerca efficace

◆ Semantica universalmente accettata

- Semantica identica per i catalogatori e per chi formula le richieste

◆ Possibilità di utilizzo in ambito internazionale

- Esistono diverse versioni nazionali

◆ Estendibilità

- Possibilità di estendere il modello per soddisfare esigenze di particolari categorie di utenti
- Si utilizzano gli “application profiles”

Dublin Core: principi generali

- ◆ **Ogni elemento è opzionale e può essere ripetuto**
 - Semplifica l'interoperabilità con altri modelli di metadati
 - Ha lo svantaggio che si possono creare record DC poco significativi
- ◆ **Ogni record di metadati DC descrive una manifestazione della risorsa**
 - Non deve esistere confusione tra la risorsa e quello che essa descrive (ad es. i metadati di una foto della Torre di Pisa dovranno descrivere la risorsa foto e non la Torre di Pisa)
- ◆ **I qualificatori sono opzionali**
 - I qualificatori fanno parte di una estensione di DC che potrebbe non essere supportata da tutti i sistemi
 - Un record DC con i qualificatori deve essere valido anche se questi vengono trascurati

Caratteristiche di Dublin Core

- ◆ **Vi sono diverse ragioni per adottare Dublin Core:**
 - I metadati Dublin Core sono facili da creare
 - Possono essere facilmente indicizzati
 - Permettono una precisione maggiore rispetto ai sistemi che utilizzano full text
 - Permettono l'interoperabilità
- ◆ **Dublin Core fornisce gli strumenti per**
 - Descrivere una risorsa (location, form etc.)
 - Fornire semplici descrizioni e permetterne l'indicizzazione
 - Individuare una risorsa
 - Ottenere ed accedere una risorsa
 - Migliorare la qualità della gestione delle risorse

Dublin Core metadata set [1/18]

- ◆ L'insieme degli elementi definiti nel Dublin Core Metadata Element Set, Version 1.1: Reference Description è presentato nella traduzione italiana, reperibile all'indirizzo <http://www.iccu.sbn.it/metadati.htm>
- ◆ Esiste un mapping tra UNIMARC e Dublin Core
 - ◆ Title
 - ◆ Creator
 - ◆ Subject
 - ◆ Description
 - ◆ Publisher
 - ◆ Contributor
 - ◆ Date
 - ◆ Type
 - ◆ Format
 - ◆ Identifier
 - ◆ Source
 - ◆ Language
 - ◆ Relation
 - ◆ Coverage
 - ◆ Rights

Dublin Core metadata set [2/18]

◆ Elemento: Titolo

Nome: Titolo

Identificatore: Title

Definizione: Un nome dato alla risorsa. Tipicamente, Title è il nome con il quale la risorsa è conosciuta

Esempi: Title = "La Divina Commedia"

Title = "La Nona Sinfonia"

Title = "Pietà"

Title = "2001: Odissea nello Spazio"

Commenti: Il titolo da solo non permette di individuare in modo univoco la risorsa

Dublin Core metadata set [3/18]

◆ Elemento: **Creatore**

Nome: Creatore

Identificatore: Creator

Definizione: Un'entità che ha la responsabilità principale della produzione del contenuto della risorsa. Esempi di creator: una persona, un'organizzazione, un servizio software.

Regole generali: I vari creator vanno riportati separatamente.

Il cognome deve precedere il nome.

Nel caso di organizzazioni strutturate, elencare le varie componenti, a partire da quella principale.

Esempi: Creator = "Alighieri, Dante"

Creator = "Internal Revenue Service. Customer Complaints Unit"

Dublin Core metadata set [4/18]

◆ Elemento: **Soggetto**

Nome: Soggetto e Parole chiave

Identificatore: Subject

Definizione: L'argomento della risorsa. Tipicamente un Soggetto viene espresso da parole chiave o frasi o codici di classificazione che descrivono l'argomento della risorsa. È buona norma selezionare il valore da un vocabolario controllato o da uno schema di classificazione.

Esempio: Subject = "Poesia"
Subject = "Sinfonia"

Commento: Non è possibile avere soggetti strutturati, quali ad esempio "software/biblioteche digitali"
In questi casi bisogna utilizzare due Soggetti

Dublin Core Metadata Set [5/18]

◆ Elemento:	Descrizione
Nome:	Descrizione
Identificatore:	Description
Definizione:	Una spiegazione del contenuto della risorsa
Commento:	Una descrizione può includere ma non è limitata solo a: un riassunto analitico, un indice, un riferimento al contenuto di una rappresentazione grafica o un testo libero del contenuto.
	<p>Molto utile in fase di ricerca. Normalmente vengono utilizzate frasi complete.</p> <p>Può essere impreciso. Inoltre classificatori diversi possono fornire descrizioni diverse della stessa risorsa.</p>
Esempio:	Description = “Questa è una foto di Piazza dei Miracoli a Pisa. A destra si vede la Torre di Pisa, al centro il Duomo ed a sinistra il Battistero”

Dublin Core Metadata Set [6/18]

- ◆ **Elemento: Editore**
 - Nome: Editore
 - Identificatore: Publisher
 - Definizione: Un'entità responsabile della produzione della risorsa, disponibile nella sua forma presente
 - Commento: Esempi di un Publisher includono una persona, un'organizzazione o un servizio. Normalmente il nome di un Publisher è usato per indicare l'entità stessa
Questo campo viene specificato per identificare l'entità che fornisce accesso alla risorsa.
 - Esempi: Publisher="Mondadori"

Dublin Core Metadata Set [7/18]

◆ Elemento: Autore di contributo subordinato

Nome: Autore di contributo subordinato

Identificatore: Contributor

Definizione: Un'entità responsabile della produzione di un contributo al contenuto della risorsa

Commento: Esempi di un Contributor includono una persona, un'organizzazione o un servizio. Normalmente il nome di un Contributor è usato per indicare l'entità stessa

Dublin Core Metadata Set [8/18]

◆ Elemento: Data

Nome: Data

Identificatore: Date

Definizione: Una data associata a un evento del ciclo di vita della risorsa. Normalmente viene associata con la data di creazione o di disponibilità della risorsa. Si raccomanda l'uso di codifiche standard per le date. Se non si conosce la data completa, si può usare solo il mese-anno o solo l'anno.

Dublin Core Metadata Set [9/18]

◆ Elemento: Tipo

Nome: Tipo di risorsa

Identificatore: Type

Definizione: La natura o il genere del contenuto della risorsa

Commento: Tipo include termini che descrivono categorie generali, funzioni, generi o livelli di aggregazione per contenuto. Si raccomanda di selezionare un valore da un vocabolario controllato (ad esempio, la lista dei “Dublin Core Types”). Per descrivere la manifestazione fisica o digitale della risorsa, si usa l’elemento Format.

Se una risorsa ha più tipi, bisogna utilizzare più campi Type.

Esempi: Type=“Text”

Type=“Sound”

Type=“Image”

Dublin Core Metadata Set [10/18]

- ◆ <http://dublincore.org/usage/terms/dcmitype/>
- ◆ **Dublin Core Types** - Dizionario dei termini usabili quali “valore” dell’elemento “TYPE” per identificare il genere di una risorsa:
 - Collection – Aggregazione di elementi.
 - Dataset – Informazione codificata in una struttura definita (lista, tabella, ...)
 - Event – Avvenimento non persistente nel tempo (conferenza, intervista, ...)
 - Image – Include anche foto, disegni, grafici, ecc.
 - Interactive Resource – richiede l’interazione con l’utente
 - Moving Image – Ad es. animazioni, film, programmi televisivi, ecc.
 - Physical Object – Oggetto tridimensionale, ad es. scultura, palazzo, ecc.
 - Service – Sistema che fornisce una o più funzioni ad un utente.
 - Software
 - Sound
 - Still image
 - Text

Dublin Core Metadata Set [11/18]

◆ Elemento:	Formato
Nome:	Formato
Identificatore:	Format
Definizione:	La manifestazione fisica o digitale della risorsa.
Commento:	<p>Normalmente Format può includere il tipo di supporto o le dimensioni della risorsa.</p> <p>Format può essere usato per determinare il software, l'hardware o altro apparato necessario alla visualizzazione o all'elaborazione della risorsa.</p> <p>Esempi di dimensioni includono grandezza e durata. Si raccomanda di selezionare un valore da un vocabolario controllato (ad esempio la lista di Internet Media Types [MIME] che definisce i formati dei supporti elettronici.</p> <p>Se vi sono più formati per una stessa risorsa, il campo Format deve essere ripetuto.</p>
Esempi:	<pre>Title="Dublin Core icon" Identifier="http://purl.org/metadata/dublin_core/images/dc2.gif" Type="Image" Format="image/gif" Format="4 kB" Subject="Saturn" Type="Image" Format="image/gif 6" Format="40 x 512 pixels" Identifier="http://www.not.iac.es/newwww/photos/images/satnot.gif"</pre>

Dublin Core Metadata Set [12/18]

INTERNET MEDIA TYPES (<http://www.iana.org/assignments/media-types/>)

Application

Audio (g722, mpeg, mpa, ...)

Image (e.g. jpeg, gif, tiff, ...)

Message

Model

Multipart

Text (e.g. rtf, sgml, richtext, xml, ...)

Video (e.g. quicktime, mpeg, H264, ...)

Dublin Core Metadata Set [13/18]

◆ Elemento: Identificatore

Nome: Identificatore della risorsa

Identificatore: Identifier

Definizione: Un riferimento univoco alla risorsa nell'ambito di un dato contesto

Commento: Si raccomanda di identificare la risorsa utilizzando un sistema formale di identificazione (ad.es. URL, URI, ISBN). Si può anche utilizzare un identificatore locale.

Esempi: Identifier="http://www.intratext.com/X/ITA0191.htm"
Identifier="ISBN:0382364759"

Dublin Core Metadata Set [14/18]

◆ Elemento: Fonte

Nome: Fonte

Identificatore: Source

Definizione: Un riferimento a una risorsa dalla quale è derivata la risorsa in oggetto

Commento: La risorsa può essere derivata dalla risorsa Source in tutto o in parte. Si raccomanda di identificare la risorsa per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito

Esempi: Source="RC607.A26W574 1996" (Call Number della versione a stampa)

Source="Immagine da pagina 32 dell'edizione pubblicata nel 1922 della Divina Commedia"

Dublin Core Metadata Set [15/18]

◆ Elemento: **Lingua**

Nome: Lingua

Identificatore: Language

Definizione: La lingua del contenuto intellettuale della risorsa. Si raccomanda l'uso di valori presi da una lista standard (RCF 3066)

Esempi: Language="it"

Language="Primarily English, with some abstracts also in French"

Dublin Core Metadata Set [16/18]

◆ Elemento: **Relazione**

Nome: Relazione

Identificatore: Relation

Definizione: Un riferimento alla risorsa correlata. Si raccomanda l'uso di un sistema di identificazione. Può indicare relazioni di tipo Parte/Tutto, Versioni, Formato derivato, citazioni, ecc.

Esempio: Title="La Divina Commedia. Inferno"
Relation="Parte 1/3"

Dublin Core Metadata Set [17/18]

◆ Elemento: Copertura

Nome: Copertura

Identificatore: Coverage

Definizione: L'estensione o scopo del contenuto della risorsa

Commento: Normalmente Coverage include la localizzazione spaziale (il nome di un luogo o le coordinate geografiche), il periodo temporale (l'indicazione di un periodo, una data o un range di date) o una giurisdizione (ad esempio il nome di un'entità amministrativa).

Si raccomanda di selezionare un valore da un vocabolario controllato (ad esempio il Thesaurus of Geographic Names [TGN]) e, se possibile, di utilizzare i nomi di luogo o i periodi di tempo piuttosto che identificatori numerici come serie di coordinate o range di date

Esempi:

Coverage="1995-1996"

Coverage="Pisa, IT"

Coverage="17th century"

Dublin Core Metadata Set [18/18]

◆ Elemento: Gestione dei diritti

Nome: Gestione dei diritti

Identificatore: Rights

Definizione: Informazione sui diritti esercitati sulla risorsa

Commento: Normalmente un elemento Rights contiene un'indicazione di gestione dei diritti sulla risorsa, o un riferimento ad un servizio che fornisce questa informazione. L'informazione sui diritti spesso comprende i diritti di proprietà intellettuale Intellectual Property Rights (IPR), Copyright e vari diritti di proprietà. Se l'elemento Rights è assente, non si può fare alcuna ipotesi sullo stato di questi o altri diritti in riferimento alla risorsa

Esempio: Rights="Accesso limitato ai membri dell'associazione"

Qualifiers [1/5]

- ◆ **Gli elementi di Dublin Core possono essere poco specifici ed ambigui in molti casi, per cui è in corso un processo di estensione basato sull'uso di *qualifiers* da associare ad ogni Dublin Core element**
- ◆ **Ogni programma che utilizza una descrizione Dublin Core deve essere in grado di ignorare i qualifiers, oppure di interpretarli**
- ◆ **Vi sono attualmente due classi di qualifiers**
 - *Element refinement*
 - ➔ *Rendono il significato di un elemento più specifico*
 - *Encoding Scheme*
 - ➔ *Semplificano l'interpretazione dell'elemento.*
 - ➔ *Possono includere vocabolari controllati o regole di elaborazione (ad es. per le date)*

Qualifiers [2/5]

- ◆ **Esempi di “element refinements”**. La lista completa si trova in <http://www.dublincore.org/documents/usageguide/qualifiers.shtml>
- ◆ **Refinement per “Title”**
 - Alternative – titolo aggiuntivo o alternativo
 - Ad es. Title = “ACM Transaction on Information Systems”
Alternative = “ACM TOIS”
- ◆ **Refinements per “Description”**
 - Table of Contents
 - Abstract
- ◆ **Refinements per “Date”**
 - Created – data di creazione della risorsa
 - Valid – data di validità della risorsa
 - Available – data di validità
 - Issued
 - Modified
 - Accepted
 - Submitted

Qualifiers [3/5]

- ◆ **Gli “Encoding Schema” sono schemi di codifica per elementi specifici di DC**
- ◆ **In particolare per Subject, Date, Type, Format, Identifier, Source, Language, Relation, Coverage sono stati definiti degli Encoding Schema**
- ◆ **Esempi:**
 - **Subject: DDC (Dewey Decimal Classification)**
 - ➔ **<http://www.oclc.org/dewey/>**
 - ➔ **Sistema di classificazione usato nel mondo librario**
 - ➔ **Fornisce un sistema per la classificazione della conoscenza**
 - ➔ **Suddivisione in 10 classi (Computers, Filosofia, Religione, ecc.)**
 - ➔ **Ogni classe è a sua volta suddivisa in sottoclassi**

Qualifiers [4/5]

◆ Esempi:

- Subject: LCC (Library of Congress Classification)
 - <http://www.loc.gov/catdir/cpsolcco/lcco.html>
 - Diverse classi (da A a Z)
 - A – General Work
 - B – Philosophy, Psychology, Religion
 - Ecc.
 - Ogni classe è a sua volta divisa in sottoclassi
- Date: DMCI Period – Codifica per specificare un periodo di tempo
 - <http://dublincore.org/documents/dcmi-period/>
 - Il periodo è specificato da [start, end, scheme, name]
 - Tutte le componenti sono opzionali
 - “start” ed “end” sono gli istanti di inizio e fine del periodo
 - “scheme” è lo schema di codifica per la rappresentazione di “start” ed “end”
 - “name” è il nome dell’intervallo di tempo
 - Esempi
 - name=The Great Depression; start=1929; end=1939;

 - name="1999 AFL Grand Final": start=1999-09-25T14:20+10:00;
end=1999-09-25T16:40+10:00; scheme=W3C-DTF;

Qualifiers [5/5]

◆ Esempi:

- Date: W3C-DTF – Regole di codifica per le date ed i tempi
 - ➔ <http://www.w3.org/TR/NOTE-datetime>
 - ➔ Sono consentite sei diverse tipologie
 - Anno: YYYY (eg 1997)
 - Anno e mese: YYYY-MM (eg 1997-07)
 - Data completa: YYYY-MM-DD (eg 1997-07-16)
 - Data completa più ore e minuti: YYYY-MM-DDThh:mmTZD (eg 1997-07-16T19:20+01:00)
 - Data completa più ore, minuti e secondi : YYYY-MM-DDThh:mm:ssTZD (eg 1997-07-16T19:20:30+01:00)
 - Data completa più ore, minuti, secondi e decimi di secondo YYYY-MM-DDThh:mm:ss.sTZD (eg 1997-07-16T19:20:30.45+01:00)
 - > dove TZD = indicatore time zone (Z o +hh:mm o -hh:mm)
 - > Il tempo viene espresso come UTC (Coordinated Universal Time) (indicato con Z) o come tempo locale più offset (+- hh:mm)
 - Esempi
 - > 1994-11-05T08:15:30-05:00 corrisponde al 5 Novembre, 1994, 8:15:30 am, US Eastern Standard Time.
 - > 1994-11-05T13:15:30Z corrisponde allo stesso istante.

Esempio di record Dublin Core

Supponiamo di voler creare un record DC per il libro di Witten, Bainbridge su Greenstone. Un possibile record Dublin Core è il seguente (si può usare l'applicazione disponibile in <http://www.lub.lu.se/cgi-bin/nmdc.pl>):

DC.Title	<i>How to Build a Digital Library</i>
DC.Creator.PersonalName	<i>Witten, Ian H.</i>
DC.Creator.PersonalName2	<i>Bainbridge, David</i>
DC.Subject	<i>Greenstone</i>
DC.Subject	Information Storage and Retrieval
DC.Subject	Documentation (CCS)
DC.Subject	H. Information Systems (CCS)
DC.Subject	H.4 Information Systems Application (CCS)
DC.Description	Il libro descrive come creare una
DC.Publisher	Morgan Kaufmann Publishers
DC.Contributor.CorporateName	University of Waikato in New Zeland
DC.Date	2003-01-01 (ISO8601)
DC.Type	Text
DC.Format	text/html (IMT)
DC.Identifier	http://pc-erato2.isti.cnr.it/savino/Greenstone-book.html
DC.Source	1-55860-790-0 (ISBN)
DC.Language	it (ISO639-1)
DC.Date.X-MetadataLastModified	2004-10-22 (ISO8601)

Rappresentazione di Dublin Core

- ◆ **Dublin Core può essere rappresentato utilizzando diversi formati**
- ◆ **Sono state sviluppate rappresentazioni sia di “simple DC” che di “qualified DC” in**
 - XML: <http://dublincore.org/documents/2003/04/02/dc-xml-guidelines/>
 - RDF: <http://dublincore.org/documents/2002/07/31/dcmes-xml/> e <http://dublincore.org/documents/dcq-rdf-xml/>

Riferimenti

- ◆ **Informazione aggiornata su Dublin Core**
<http://dublincore.org/index.shtml>
- ◆ **La documentazione aggiornata si trova a partire dall'indirizzo:** <http://dublincore.org/usage/>
- ◆ **Diane Hillman, Using Dublin Core,**
<http://dublincore.org/documents/usageguide/>
- ◆ **Dublin Core Metadata Template**
<http://www.lub.lu.se/cgi-bin/nmdc.pl>
- ◆ **Crea dublin core da pagine web**
<http://www.ukoln.ac.uk/metadata/dcdot/>