

## Capitolo 4

# Metadati

### 4.1 Introduzione

Una Biblioteca Digitale (BD) contiene un insieme di documenti, caratterizzati dal loro contenuto – cioè dall’informazione presente nel documento – dal modo in cui il contenuto è strutturato e dalle informazioni associate ad ogni documento ma non direttamente presenti nel documento stesso.

Per quanto riguarda il contenuto, possiamo avere Biblioteche Digitali che contengono documenti di vario tipo, come ad esempio testo, immagini, filmati, oppure una combinazione di diversi tipi di informazione. Per quanto riguarda la struttura dei documenti, normalmente ci si riferisce alla possibilità di suddividere il contenuto in diverse parti, eventualmente a loro volta composte di sottoparti. Un ben noto esempio di documento strutturato sono i libri, che contengono un titolo, una lista di autori, una prefazione, un insieme di capitoli, a loro volta suddivisi in sezioni, ecc. Si possono comunque immaginare altri esempi di documenti strutturati: ad esempio, una presentazione può consistere di una introduzione, un insieme di parti per i diversi argomenti trattati, ed una conclusione. Ogni lucido della presentazione avrà una struttura variabile, costituita ad esempio da una descrizione testuale ed una o più immagini. Infine, ogni documento può essere descritto da informazioni non direttamente presenti al proprio interno, ma tuttavia necessarie per caratterizzarlo. Si pensi, ad esempio ad una Biblioteca Digitale contenente le foto di quadri. Ogni singola foto ha bisogno di essere descritta tramite informazioni aggiuntive quali l’autore del quadro, il titolo, la data di produzione, il luogo nel quale viene conservato, ecc.

Per poter operare agevolmente sui documenti contenuti nella Biblioteca Digitale dobbiamo fornire una descrizione precisa di tutte le caratteristiche

dei documenti in essa contenuti. Bisogna comunque considerare che la descrizione dei documenti è fortemente dipendente dal tipo di funzionalità che la BD deve fornire. Se ad esempio vogliamo dare una descrizione del contenuto dei documenti per poter effettuare delle ricerche, avremo bisogno di fornire il titolo, l'autore, ecc. Se invece vogliamo controllare e monitorare lo stato di conservazione dei documenti dovremo fornire una descrizione dei materiali dei quali è composto, delle operazioni di restauro eventualmente effettuate, dei fenomeni di degrado rilevati, ecc.

In generale, la *descrizione* del documento sarà fornita da un insieme di diverse tipologie di informazione indicate con il nome generico di *metadati*.

Non è semplice dare una definizione precisa e completa del termine "metadato". Una definizione semplice, anche se abbastanza generica, è la seguente:

*I metadati sono informazioni **strutturate** relative ai **dati**, interpretabili da parte di un computer.*

In questa definizione gli aspetti essenziali sono dati dall'indicazione che i metadati forniscono informazioni relative ai dati, che queste informazioni hanno una struttura ben definita e che, grazie a questa strutturazione, questa informazione può essere interpretata ed utilizzata dal programma di un computer.

Una definizione più precisa può essere data operazionalmente, individuando le tipologie di risorse<sup>1</sup> alle quali usualmente vengono associati dei metadati, le tipologie di informazione contenute nei metadati e le operazioni che possono essere effettuate sui metadati.

*I metadati sono informazioni **strutturate** appartenenti a diverse **tipologie**, che vengono utilizzate per effettuare un insieme di **operazioni** su un insieme di **risorse***

In generale, i metadati possono essere associati a qualunque *risorsa*, sia essa digitale, fisica, o astratta. Quindi, per riportare solo alcuni esempi, possiamo associare metadati a libri, record bibliografici, documenti HTML, oggetti museali, siti Web, collezioni di oggetti, servizi, luoghi, persone, istituzioni, opere astratte, concetti, eventi, ecc.

Per quanto riguarda invece le *tipologie* di informazione contenute nei metadati, queste possono essere distinte in diverse categorie, delle quali nel seguito riportiamo alcuni esempi:

---

<sup>1</sup>Nel seguito al posto del termine *documento* useremo il termine più generale *risorsa*

**Record bibliografici.** In questo caso i metadati forniscono informazione altamente strutturata degli elementi essenziali degli oggetti contenuti nella Biblioteca Digitale. Questa informazione spesso non è contenuta direttamente nella risorsa: per esempio, la posizione di un libro in una biblioteca non è relativa al contenuto del libro.

**Sommario (Abstract).** In questo caso il metadato contiene informazione meno strutturata ed organizzata. Questo tipo di informazione è normalmente legato al contenuto della risorsa: per esempio, l'abstract di un libro ne descrive in modo sintetico il contenuto. L'abstract può essere prodotto sia manualmente da un operatore specializzato che da strumenti automatici.

**Descrizione del materiale archiviato.** Molto spesso vengono forniti metadati che descrivono il contenuto delle risorse. Ad esempio, possiamo fornire una serie di categorie alle quali la risorse appartiene, oppure una lista di parole chiave.

**Descrizione del contenuto digitale delle risorse.** Le risorse possono essere di diverso tipo quali ad esempio immagini, testo, audio, video. Una tipologia di metadati sono i descrittori del contenuto digitale di ogni risorsa. Ad esempio, per un'immagine potremo avere una descrizione dei colori in essa contenuti, per un testo avremo la lista delle parole che lo compongono, per un video la lista delle diverse scene.

**Descrizione delle caratteristiche fisiche della risorsa.** Un'informazione molto utile, principalmente per permettere la gestione, visualizzazione e manipolazione delle diverse risorse riguarda l'indicazione delle caratteristiche fisiche della risorsa. Ad esempio, per un testo sarà necessario specificare che caratteri vengono usati, per un'immagine il formato (bmp, Jpeg, ecc.) utilizzato per la sua rappresentazione, ecc.

**Descrizione dello stato di conservazione.** Spesso una risorsa (sia digitale che fisica) può essere soggetta a degrado con conseguente perdita di informazione, per cui è necessario conoscere quando la risorsa è stata creata, quanto è stata eventualmente restaurata, quali strumenti sono stati utilizzati per la generazione ed il restauro e, nel caso di risorse digitali, quali strumenti possono essere usati per la presentazione e per la visualizzazione.

**Descrizione dei diritti di utilizzo.** I metadati possono essere utilizzati per specificare quali utenti o quali categorie di utenti possono accedere

ed in che forma (ad esempio in sola lettura o anche in modifica) ad una risorsa o ad una sua parte.

**Diritti di proprietà.** Una risorsa digitale può essere di proprietà esclusiva o parziale di una persona o di una organizzazione. Inoltre, questa proprietà può essere soggetta a limitazioni, per esempio nella durata.

Infine, diamo una lista delle principali *operazioni* che possono essere effettuate sulle risorse tramite l'utilizzo dei metadati.

**Ricerca di informazioni** (*Resource disclosure & discovery*). La possibilità di effettuare la ricerca delle risorse contenute nella Biblioteca Digitale è uno degli utilizzi principali dei metadati.

**Gestione delle risorse** (*Resource management*). Questa funzionalità è abbastanza ampia, ed include tutta una serie di funzioni che vanno dalla semplice gestione della risorsa (ad esempio è possibile specificare dove è stata archiviata, quando è stato fatto il backup, ecc.), alla gestione della conservazione della risorsa, alla possibilità di accedervi nel futuro, fino alla gestione delle modalità di pagamento per l'accesso.

**Gestione dei diritti di proprietà** (*Intellectual property rights management*). In molti casi la risorsa o parti della risorsa sono di proprietà di persone o di un'organizzazione. I metadati possono essere utilizzati per evitare usi non consentiti della risorsa e per individuarne il proprietario e gli utilizzatori autorizzati.

**Autenticazione ed autorizzazione.** Questa funzione è strettamente legata alla precedente, poichè ogni controllo della proprietà intellettuale richiede che gli utenti si debbano autenticare e debbano essere autorizzati ad utilizzare il sistema e ad accedere alle risorse disponibili. Questa funzione permette anche di diversificare gli utenti in base alle tipologie di risorse alle quali possono accedere.

**Personalizzazione e localizzazione dei servizi.** Spesso può essere utile fornire servizi specifici ad ogni utente. Inoltre, può essere opportuno che il sistema sia adattabile alle esigenze di ogni utente: ad esempio, per un utente che è interessato in modo particolare ad argomenti di tipo storico, in risposta ad una interrogazione, il sistema potrà dare maggiore importanza a documenti storici rispetto ad altri.

Infine, vediamo come possono essere descritti i metadati associati ad una risorsa. Ogni *tipologia* di informazione associata ai metadati sarà descritta tramite un insieme di *proprietà*, ognuna rappresentata da un *elemento* (*metadata element*). Ogni *elemento*, come vedremo nel seguito, può avere una diversa e ben definita rappresentazione. Un insieme coerente di *elementi di metadati* costituisce un *insieme di elementi di metadati* (*metadata element set*), che viene creato per uno scopo specifico – ad esempio, per effettuare una specifica operazione su un dato tipo di risorsa – ed è normalmente definito come una singola unità.

#### 4.1.1 Vari tipi di metadati

Come visto, i metadati sono caratterizzati dal tipo di risorsa che descrivono e dalle operazioni che la BD deve fornire per quella data risorsa. Su questa base, si possono individuare tre classi principali di metadati:

**Metadati descrittivi**<sup>2</sup> Questi metadati vengono utilizzati per descrivere le risorse che costituiscono la Biblioteca Digitale, principalmente allo scopo di semplificare e migliorare la ricerca delle risorse stesse da parte degli utenti. Usualmente, questi metadati vengono organizzati in cataloghi, oppure vengono generati degli indici usati per la ricerca di informazione. Esempi tipici di metadati descrittivi sono, nel caso di biblioteche contenenti documenti, il *Titolo*, l'*Autore*, la *Data di pubblicazione*, ecc. Ovviamente, avremo bisogno di metadati descrittivi diversi per le diverse tipologie di risorse che vogliamo descrivere e per le diverse funzioni di ricerca che vogliamo che la biblioteca digitale fornisca.

**Metadati strutturali.** Come dice il nome, questi metadati permettono di rappresentare la struttura delle risorse contenute nella biblioteca digitale. Inoltre, i metadati strutturali permettono di descrivere le relazioni esistenti tra risorse diverse o tra parti di una stessa risorsa. Per rimanere all'esempio più usuale di biblioteca digitale di documenti, i metadati strutturali possono descrivere i riferimenti tra documenti diversi (come ad esempio le citazioni), oppure possono descrivere la struttura del documento, costituito di *capitoli*, *sezioni*, ecc.

**Metadati amministrativi.** I metadati amministrativi forniscono una descrizione delle risorse che permette di effettuare le operazioni di ges-

---

<sup>2</sup>Spesso i metadati descrittivi sono indicati come metadati per *resource discovery*, essendo questo il loro principale utilizzo

tione della biblioteca – ad esempio le attività di manutenzione – e per il controllo degli accessi e della sicurezza. Per esempio, potremmo avere una indicazione delle tipologie di utenti che possono accedere alla risorsa (o a parti di essa), oppure potremmo avere una descrizione degli ultimi back-up effettuati e di dove si trovino le copie di back-up effettuate. Nell'ambito dei metadati amministrativi possiamo inoltre includere la categoria dei metadati utilizzati per la conservazione delle risorse digitali, o *digital preservation* come viene normalmente denominata.

Supponiamo, ad esempio, di voler fornire dei metadati relativi a questo libro. I metadati descrittivi potrebbero essere del tipo:

<b>Titolo:</b>	Appunti del corso di Biblioteche Digitali
<b>Autore:</b>	Claudio Gennaro
<b>Autore:</b>	Pasquale Savino
<b>Data di pubblicazione:</b>	Gennaio 2005
<b>Parola chiave:</b>	Biblioteche Digitali
<b>Parola chiave:</b>	Metadati

I metadati strutturali potrebbero specificare che il libro è suddiviso in capitoli, paragrafi e sottoparagrafi e che è legato ai lucidi delle lezioni ed al software che viene utilizzato per le esercitazioni. Potremmo quindi avere i seguenti metadati strutturali:

#### Capitolo

**Titolo:** Introduzione alle Biblioteche Digitali

##### Sezione

**Titolo:** Definizione di Biblioteca Digitale

#### Capitolo

**Titolo:** XML

##### Sezione

**Titolo:** Introduzione

#### Link

**Tipo:** Presentazione ppt

**Titolo:** Lucidi del corso

**Riferimento:** indirizzo web

#### Link

**Tipo:** Software

**Titolo:** Greenstone

**Riferimento:** indirizzo web

Si noti che nell'esempio precedente i titoli dei capitoli, delle sezioni e dei link sono metadati descrittivi.

I metadati amministrativi specificano invece gli utenti che sono proprietari della risorsa, la lista di distribuzione e la data dell'ultimo backup effettuato.

<b>Proprietà:</b>	Claudio Gennaro
<b>Proprietà:</b>	Pasquale Savino
<b>Distribuzione:</b>	Studenti I.U.
<b>Data backup:</b>	20050410

Nel seguito ci occuperemo principalmente dei metadati descrittivi, essenziali in una biblioteca digitale per fornire alcune delle principali funzionalità che essa deve fornire. Vediamo sinteticamente quali sono le principali funzionalità richieste dagli utenti:

**Ricerca.** Gli utenti sono interessati ad individuare le risorse in accordo a determinati criteri di ricerca. Ad esempio, una tipica richiesta potrebbe essere: *"Trova tutte le opere di Dante che parlano di Beatrice"*. Senza l'utilizzo di metadati adeguati, questa richiesta può risultare difficile da soddisfare, oppure può produrre risultati diversi da quelli desiderati.

**Identificare una specifica risorsa.** Gli utenti sono spesso interessati a distinguere e a differenziare diverse risorse che risultano simili tra loro. Ad esempio, dopo aver individuato tutti i libri che parlano delle Biblioteche Digitali, siamo interessati a selezionarne alcuni sulla base di specifici criteri, che possono essere la lingua, particolari argomenti trattati, il modo in cui vengono descritti gli argomenti, l'autore, ecc. Anche in questo caso l'utilizzo di metadati descrittivi può semplificare l'operazione.

**Localizzare la risorsa.** Dopo aver individuato le risorse che sono rilevanti alle proprie esigenze, l'utente deve localizzare la posizione fisica della risorsa per potervi accedere.

**Ottenere/accedere alla risorsa.** La fase successiva consiste nell'ottenere la risorsa, in modo da poterla visualizzare ed eventualmente modificare.

**Utilizzare la risorsa.** La fase conclusiva consiste nella possibilità, da parte dell'utente di utilizzare la risorsa in modo completo, visualizzandola, modificandola, ecc.

I metadati descrittivi possono essere suddivisi sulla base delle modalità di generazione, che portano ad avere:

**Metadati estratti automaticamente.** Appartengono a questa categoria il testo dei documenti e le feature delle immagini (ad esempio, la distribuzione di colore, la tessitura, la forma degli oggetti). L'estrazione automatica avviene utilizzando dei moduli software appositamente sviluppati.

**Metadati inseriti manualmente.** Questa categoria include tutti i metadati che non appartengono alla precedente. Sarà necessario l'intervento umano per fornire i valori dei metadati in tutti i casi in cui la complessità dell'estrazione automatica è superiore alle capacità degli strumenti di calcolo, e nei casi in cui l'informazione da inserire non è direttamente presente nella risorsa. Questo ultimo caso è abbastanza frequente, come evidenziato per nell'esempio riportato all'inizio del capitolo relativo alle foto dei quadri.

Sulla base della significatività e complessità dei metadati possiamo distinguere due diversi gruppi:

**Forme semanticamente semplici.** Forniscono una descrizione generica della risorsa, per cui possono essere utilizzate per diverse tipologie di risorse, anche se non permettono un dettaglio sufficiente nei casi in cui è necessario formulare interrogazioni su aspetti particolari. Questi metadati sono usualmente immessi manualmente dagli utenti, attraverso degli strumenti software che agevolano l'immissione dei dati. Questa operazione non richiede particolari competenze, trattandosi di metadati poco specifici.

**Forme semanticamente complesse.** In alcuni casi abbiamo bisogno di descrivere le risorse in modo molto dettagliato, oppure dobbiamo fornire una descrizione di aspetti particolari, oppure dobbiamo descrivere delle risorse particolari (ad esempio documenti Audio/Video). In questi casi è necessario utilizzare dei modelli di metadati complessi, che permettono di descrivere in tutta la sua ricchezza il contenuto della risorsa. Per esempio, nel caso di un documento Audio/Video potremmo voler individuare le diverse scene e, per ognuna, fornire una descrizione



dei personaggi, dell'ambiente, ecc. Nella maggior parte dei casi questi metadati vengono immessi manualmente dagli utenti, utilizzando strumenti software specifici. Inoltre, la classificazione deve essere effettuata da personale specializzato, che sia in grado di descrivere la risorsa in modo completo e corretto.

Nella descrizione precedente abbiamo caratterizzato i metadati a seconda che vengano estratti automaticamente o che vengano inseriti manualmente da un utente che svolge il compito di catalogatore. Al riguardo possiamo fare alcune semplici considerazioni:

- La catalogazione automatica è preferibile rispetto a quella manuale, poichè permette una riduzione dei costi e dei tempi e perchè consente di ottenere catalogazioni univoche. Quest'ultimo aspetto è particolarmente importante poichè una catalogazione oggettiva è necessaria per ottenere una migliore qualità dei risultati della ricerca<sup>3</sup>. Infatti, quando si catalogano le risorse manualmente, due diversi catalogatori possono fornire valori di metadati diversi per la stessa risorsa; questo implica che quando si cercherà quella particolare risorsa, la ricerca darà risultati diversi a seconda di chi ha effettuato la catalogazione.
- Le considerazioni precedenti a favore della catalogazione automatica non tengono conto della necessità di fornire, in molti casi, metadati che non possono essere estratti da un programma software, o per la complessità del compito oppure perchè il valore del metadato non è direttamente presente nella risorsa.

#### 4.1.2 Associazione dei metadati alle risorse

Prima di passare alla descrizione di alcuni dei più comuni modelli di metadati utilizzati per il resource discovery, consideriamo il problema di come associare i metadati ad una specifica risorsa. In generale possiamo considerare tre diverse possibilità, ognuna con i propri vantaggi e svantaggi:

1. I metadati possono essere contenuti all'interno della risorsa. Per esempio, nel caso la risorsa sia un documento HTML, l'uso del *tag META*

---

<sup>3</sup>Il problema della *qualità* dei risultati di una ricerca è centrale nel settore dell'Information Retrieval. In modo sintetico possiamo dire che ogni per ricerca di informazioni esiste un insieme ideale dei risultati che soddisfano le richieste dell'utente. L'obiettivo è quello di avvicinarci quanto più possibile a questo risultato ideale, riducendo sia il numero di documenti significativi non ritrovati che il numero di documenti non significativi ritrovati (*falsi positivi*)

permette di specificare valori di metadati che non influenzano il modo in cui il documento viene visualizzato.

2. Il record dei metadati è un oggetto separato dalla risorsa che contiene al suo interno un identificatore del metadato. Per esempio, se la risorsa è un documento HTML, possiamo usare il *tag link* per specificare un puntatore ad un record contenente i metadati del documento.
3. Il record dei metadati è un oggetto separato dalla risorsa e contiene al suo interno un identificatore della risorsa. In questo caso non si modifica in alcun modo la risorsa, per cui l'associazione può essere fatta per risorse di ogni tipo.

La prima e la seconda soluzione richiedono che le risorse offrano uno strumento per descrivere i metadati (come ad esempio i *tag meta* e *link* di HTML). Questa condizione non è soddisfatta in generale. Inoltre, anche nel caso che sia possibile rappresentare il valore del metadato, sarà necessario che il creatore di metadati abbia i diritti per la modifica della risorsa e che sia fornito un programma specifico per estrarre questi valori. La seconda soluzione richiede inoltre che venga garantito il vincolo di integrità dei record di metadati, per cui non sia permesso cancellare tale record senza aver modificato la risorsa. Tutte queste difficoltà sono superate se si adotta la soluzione di mantenere una completa indipendenza tra il record di metadati e la risorsa. In questo caso non è richiesta la modifica della risorsa ed il record di metadati può essere cancellato o modificato senza difficoltà. Inoltre, si possono facilmente associare più record di metadati alla stessa risorsa (ad esempio metadati descrittivi e metadati amministrativi). Nella pratica questa è la soluzione che viene adottata più di frequente.

### 4.1.3 Modelli di metadati

Abbiamo visto che i metadati vengono utilizzati per scopi diversi e per descrivere diverse tipologie di risorse. Risulta quindi naturale che a seconda dell'uso e della risorsa che si vuole descrivere vengano utilizzati modelli di rappresentazione diversi. Nel seguito ci occuperemo di alcuni tra i principali modelli di metadati descrittivi, utilizzati per la descrizione di record bibliografici (il modello MARC) e di risorse accedibili dal Web (il modello Dublin Core). Nei prossimi due capitoli illustreremo due modelli per dati multimediali, il modello MPEG-7 ed il modello ECHO, specifico per documenti audio/video.

## 4.2 MARC

L'uso di schede descrittive delle opere contenute nella biblioteca e di indici che semplificano la ricerca delle opere stesse è stata adottata da lungo tempo nelle Biblioteche tradizionali. L'avvento dei calcolatori ed il loro utilizzo all'interno delle biblioteche ha portato alla creazione di schede in formato elettronico contenenti la descrizione delle opere e di indici, gestiti dal calcolatore ed interrogabili utilizzando gli strumenti propri delle basi di dati. Le prime esperienze in tal senso sono state condotte indipendentemente presso diverse biblioteche nazionali, ed ogni organizzazione ha definito dei formati di schede descrittive incompatibili tra loro. Questo ha posto il problema di definire un formato record standard, che fosse utilizzabile da più biblioteche, in modo da semplificare e velocizzare sia il processo di immissione dei dati che per poter standardizzare gli strumenti di ricerca delle informazioni. A questo scopo è stato sviluppato, alla fine degli anni '60, presso la Library of Congress, il modello MARC (Betty FurrieMACHINE Readable Cataloging)<sup>4,5</sup>. MARC, inizialmente sviluppato per il solo mercato statunitense, è stato successivamente adottato, con diverse varianti, da altri organismi bibliotecari nazionali; sono quindi nati l'UKMARC (per la Gran Bretagna), il CANMARC (per il Canada), l'ANNAMARC (per l'Italia), ecc. L'ampia diffusione del formato MARC ed il proliferare di formati simili ma leggermente diversi tra loro, ha posto il problema della definizione di uno standard internazionalmente riconosciuto: è quindi nato UNIMARC<sup>6</sup>, che può anche essere utilizzato per lo scambio di dati tra formati MARC diversi.

Formati specifici sono stati definiti per cinque diversi tipi di dati: bibliografici, holdings, authority, classificazione, community information. Questi diversi formati utilizzano tutti le stesse regole di composizione del record MARC, ma differiscono per i codici utilizzati per ogni elemento del record.

Nel seguito descriveremo la struttura del record MARC21, ed alcuni esempi di record bibliografici MARC21. Il MARC21 è derivato dall'unificazione ed armonizzazione – iniziata nel 1994 e conclusasi nel 1999 – tra il modello MARC degli Stati Uniti (USMARC) e del Canada (CAN/MARC).

---

<sup>4</sup>Betty Furrie, Understanding MARC. Bibliographic, Machine-readable cataloging, <http://www.loc.gov/marc/umb/>

<sup>5</sup><http://www.loc.gov/marc/>

<sup>6</sup><http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>

LC Control Number: 95768890  
 Type of Material: Music Sound  
 Recording Main Title: The Beatles and Frank Ifield on stage [sound recording].  
 Published/Created: [S.l.] : Vee-Jay, [1964?]  
 Related Names: Ifield, Frank. prf  
                   Beatles. prf  
 Description: 1 sound disc : analog, 33 1/3 rpm ; 12 in.  
 Publisher Number: VJ-1085 Vee-Jay  
 Contents: Please, please me -- Any time --  
           Lovesick blues -- I'm smiling now --  
           Nobody's darling -- From me to you --  
           I remember you -- Ask me why --  
           Thank you girl -- The wayward wind --  
           Unchained melody -- I listen to my heart.  
 Notes: Brief record.  
 Performed/spoken by: Performed by the Beatles and Frank Ifield.  
 Subjects: Popular music--1961-1970.  
 LC Classification: Vee-Jay VJ-1085  
 Other System No.: (OCoLC)33448757

Figura 4.1: Esempio di record MARC di un disco dei Beatles

#### 4.2.1 Descrizione generale

Come accennato, l'acronimo MARC sta per "Machine readable cataloging", quindi un record di informazione che può essere letto, interpretato e manipolato da un computer.

L'informazione contenuta in ogni record MARC dipende dal particolare oggetto della biblioteca che si vuole descrivere. In ogni caso, ogni record MARC contiene (i) una *descrizione* dell'oggetto, eventualmente con una indicazione che si tratta della descrizione dell'oggetto principale di un gruppo di oggetti, o degli oggetti componenti (ad es. per una serie di opere), (ii) un insieme di *soggetti* selezionati da una lista predefinita, (iii) ed un *numero di collocazione* che indica dove si trova fisicamente l'oggetto (lo scaffale della biblioteca).

La struttura dei record MARC è molto precisa, in modo da permettere l'interpretazione da parte di programmi software. Ogni elemento ha uno specifico identificatore numerico che lo individua in modo univoco, ed una precisa struttura. Nella figura 4.1 è riportato l'esempio di un record

bibliografico della Library of Congress, mentre in figura 4.2 è riportato il corrispondente record in formato MARC.

#### 4.2.2 Struttura dei record MARC

La descrizione completa e dettagliata dei record MARC è al di là degli obiettivi di questo testo. Nel seguito cercheremo di evidenziare gli elementi principali della struttura dei record MARC, mentre rimandiamo alla documentazione fornita dalla Library of Congress per una trattazione più dettagliata<sup>7</sup>.

Ogni record MARC è composto di una intestazione, un insieme di campi di controllo ed un insieme di campi con i dati. Più precisamente la struttura è la seguente:

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

dove

- `<leader>` è un campo a lunghezza fissa che contiene informazione generale sul record,
- `<directory>` è un indice alla posizione dei campi di controllo e dati (`<control-field>` e `<data-field>`) all'interno del record
- `<control-number-field>` contiene un singolo carattere di controllo specifico dell'organizzazione che ha generato il record MARC
- `<control-field>` è un campo, che può essere ripetuto più volte, che contiene informazione di controllo specifica del record. Questi campi iniziano con un tag di tre cifre, avente le prime due poste a zero (quindi hanno la forma 00X).
- `<data-field>` è un campo, anch'esso ripetibile, che contiene i valori di metadati
- FT è un carattere di controllo che indica la terminazione di un campo
- RT è un carattere di controllo che indica la terminazione del record

---

<sup>7</sup><http://www.loc.gov/marc/>

LC Control Number: 95768890  
 000 01128cjm 22003137a 450  
 001 5745525  
 005 19970714083147.0  
 007 sdubumennmplue  
 008 951109s1964 xx uun eng  
 035 ## |9 (DLC) 95768890  
 906 ## \$a 7 \$b cbc \$c orignew \$d 3 \$e ncip \$f 19 \$g y-genmusic  
 955 ## \$a vn31  
 010 ## \$a 95768890  
 028 02 \$a VJ-1085 \$b Vee-Jay  
 035 ## \$a (OCoLC)33448757  
 040 ## \$a DLC \$c DLC \$d DLC  
 050 00 \$a Vee-Jay VJ-1085  
 245 04 \$a The Beatles and Frank Ifield on stage \$h [sound recording].  
 260 ## \$a [S.l.] : \$b Vee-Jay, \$c [1964?]  
 300 ## \$a 1 sound disc : \$b analog, 33 1/3 rpm ; \$c 12 in.  
 511 0# \$a Performed by the Beatles and Frank Ifield.  
 505 0# \$a Please, please me -- Any time --  
           Lovesick blues -- I'm smiling now --  
           Nobody's darling -- From me to you --  
           I remember you -- Ask me why --  
           Thank you girl -- The wayward wind --  
           Unchained melody -- I listen to my heart.  
 500 ## \$a Brief record.  
 650 #0 \$a Popular music \$y 1961-1970.  
 700 1# \$a Ifield, Frank. \$4 prf  
 710 2# \$a Beatles. \$4 prf  
 953 ## \$a TA28  
 985 ## \$a 9511  
 991 ## \$b c-RecSound \$h Vee-Jay VJ-1085 \$w MUSIC

Figura 4.2: Esempio di record MARC

### Il campo <leader>

Il campo <leader> è composto di 24 byte di caratteri ASCII con la seguente struttura:

Posizione	Nome	Descrizione
0 – 4	record-length	Lunghezza del record
5	record-status	Un singolo carattere ASCII che indica lo stato del record (ad es. nuovo, modificato, ecc.)
6	type-of-record	Carattere ASCII che specifica le caratteristiche del record
7 – 8	implementation-defined	Riservato a specifiche implementazioni
9	character-coding-schema	Specifica lo schema di codifica dei caratteri utilizzato
10	indicator-count	Specifica il numero di indicatori presenti in ogni campo dati
11	subfield-code-length	Specifica la somma della lunghezza del delimitatore e dell'identificatore dei dati
12 – 16	base-address-of-data	Specifica la posizione del primo carattere del campo variabile del record
17 – 19	implementation-defined	Riservato a specifiche implementazioni
20 – 23	entry-map	Quattro caratteri che specificano la struttura del del campo directory

### Il campo <directory>

Il campo <directory>, costituito da un insieme di elementi, di lunghezza fissa, uno per ogni campo variabile del record, fornisce informazioni sui campi variabili del record. La struttura di ogni elemento della directory è la seguente:

Posizione	Nome	Descrizione
0 – 2	tag	Caratteri utilizzati per identificare o per assegnare una etichetta al campo
3 – 6	length-of-field	Lunghezza del campo
7 – 11	starting-character-position	Posizione iniziale del campo

### Il campo <data-field>

Il campo <data-field> contiene i dati del record MARC. 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.

Schematicamente ogni campo <data-field> è costituito da una sequenza di elementi del tipo:

```
<field-tag> [<indicatore-1>] [<indicatore-2>] [<delimiter>]
<subfield-tag>,<subfield-value>]+
```

dove <field-tag> è rappresentato da un codice di 3 cifre, <indicatore-1> ed <indicatore-2> sono campi opzionali che specificano come deve essere interpretato il valore associato al <field-tag>. Vi possono essere uno o più sotto-campi, ognuno composto da un delimitatore (<delimiter>), da un identificatore del sottocampo (<subfield-tag>) e dal valore corrispondente (<subfield-value>). Il delimitatore verrà rappresentato nel seguito con il simbolo \$. Il <subfield-tag> è costituito da una lettera minuscola. Consideriamo il seguente esempio:

```
300 ## $a 675p.
      $b ill.
      $c 24 cm.
```

che si riferisce al campo descrizione fisica di un libro che ha il <field-tag> 300 ed è costituito da tre sotto-campi, individuati rispettivamente dai <subfield-tag> *a*, *b*, *c* (\$ è il delimitatore del sotto-campo) che corrispondono al *numero di pagine*, alle *informazioni sulle illustrazioni*, e alle *dimensioni fisiche del libro*.



```

01128cjm 22003137a 4500001000800000005001700008007001500025
008004100040035002100081906004500102955000900147010001700156
028002100173035002100194040001800215050002000233245006200253
260003200315300004900347511004700396505022400443500001800667
650003000685700002400715710001800739953000900757985000900766
991003900775^5745525^19970714083147.0^sdubumennmplue^951109s
1964 xx uun eng ^##$9(DLC) 95768890^##$a7
$bcbc$corignew$d3$encip$f19$gy-genmusic^##$avn31^##$a 9576
8890 ^02$aVJ-1085$bVee-Jay^##$(0CoLC)33448757 ^##$aDLC$cDLC
$dDLC^00$aVee-Jay VJ-1085^04$aThe Beatles and Frank Ifield o
n stageh[sound recording].^##$a[S.l.] :$bVee-Jay,$c[1964?]^
##$a1 sound disc :$banalog, 33 1/3 rpm ;$c12 in.^0#$aPerform
ed by the Beatles and Frank Ifield.^0#$aPlease, please me --
Any time -- Lovesick blues -- I'm smiling now -- Nobody's d
arling -- From me to you -- I remember you -- Ask me why --
Thank you girl -- The wayward wind -- Unchained melody --
I listen to my heart.^##$aBrief record.^#0$aPopular music$y
1961-1970.^1#$aIfield, Frank.$4prf^2#$aBeatles.$4prf^##$aTA
28^##$a9511^##$bc-RecSound$hVee-Jay VJ-1085$wMUSIC^

```

Figura 4.3: Esempio di record MARC

### 4.2.3 Un esempio di record MARC

Il record MARC delle figure 4.1 ed 4.2 è riportato nella figura 4.3. Il record MARC completo contiene i campi <leader>, <directory>, i campi di controllo ed i campi dati.

Nella figura 4.4 viene mostrata la sola parte relativa ai campi <leader> e <directory> e vengono evidenziati i diversi elementi che li compongono<sup>8</sup>.

Dall'esempio precedente si deduce che il tag 001 ha per valore ^5745525 mentre il tag 260, ad esempio, corrisponde a ^##\$a[S.l.] :\$bVee-Jay,\$c[1964?].

### 4.2.4 I field-tags

I <field-tag> sono codici di tre cifre che individuano in modo univoco il campo (*field*). I <field-tag> sono suddivisi in 9 categorie, individuate dalla prima delle tre cifre; le categorie sono diverse per i vari tipi di dati che

<sup>8</sup>Si ricordi che ogni elemento della *directory* è composto di tre caratteri corrispondenti al *tag*, quattro caratteri che indicano la lunghezza del campo e cinque caratteri che corrispondono alla posizione di inizio del campo

```
01128cjm 22003137a 4500 leader
001000800000 tag 001 lunghezza 0008 inizio 00000
005001700008 tag 005 lunghezza 0017 inizio 00008
007001500025 tag 007 lunghezza 0015 inizio 00025
008004100040 ecc.
035002100081
906004500102
955000900147
010001700156
028002100173
035002100194
040001800215
050002000233
245006200253
260003200315
300004900347
511004700396
505022400443
500001800667
650003000685
700002400715
710001800739
953000900757
985000900766
991003900775
```

Figura 4.4: Campi <leader> e <directory> di un record MARC

vengono gestiti in MARC (i dati bibliografici, le holding, ecc.). Nel caso che a noi interessa maggiormente, quello dei record bibliografici, si distinguono le seguenti categorie:

- 0XX** Informazione di controllo, numeri, codici. Può contenere informazioni quali l'identificatore dell'organizzazione che ha creato il record, la data dell'ultima transazione effettuata sul record, informazione sulle caratteristiche fisiche dell'oggetto (ad es. il numero dei colori per un'immagine, il formato di presentazione per un video, ecc.)
- 1XX** Parte principale. Questo tag può corrispondere a nomi di persone, di organizzazioni, di luoghi, ecc., utilizzati come punti principali di accesso al record bibliografico.
- 20X–24X** Titoli, edizione, statement di responsabilità, informazione sulla pubblicazione.
- 250–270** Edizione, informazioni sulla pubblicazione, stampa, distribuzione, indirizzo, ecc.
- 3XX** Descrizione fisica (ad esempio, il numero di pagine, di volumi, oppure le dimensioni), frequenza di pubblicazione, prezzo, data di pubblicazione.
- 4XX** Serie. Contiene informazioni specifiche per le opere in serie (per es. i diversi numeri di una rivista o i volume di un'enciclopedia). In questo caso si può specificare il titolo ed il volume della parte.
- 50X–53X** Note. Sono incluse in questa categoria varie tipologie di note, quali ad esempio le note di una dissertazione, le note bibliografiche, le citazioni, la data ed il luogo di un evento, ecc.
- 53X–58X** Contiene note più specifiche di quelle della categoria precedente. Esempi sono le note sui diritti di utilizzo e riproduzione, la lingua o le lingue utilizzate, i diritti di proprietà e di custodia, ecc.
- 6XX** Soggetto. Si tratta di elementi che forniscono punti di accesso addizionali ai record. I soggetti si distinguono in varie tipologie quali *nomi di persona*, *nomi geografici*, *termine cronologico*, *genere*, ecc.
- 70X–75X** Altre entries diverse dai campi soggetto, serie o titolo. Si distinguono tipologie quali *nomi di persona*, *nomi di luoghi*, *identificazioni tassonomiche*, ecc.

**76X–78X** Collegamenti (link) con altri elementi. I campi appartenenti a questa categoria contengono informazione che permette di identificare altri elementi bibliografici. I diversi campi specificano le relazioni tra l'elemento descritto e quello riferito. Questi campi sono usati principalmente per generare, in fase di visualizzazione del record, delle note che descrivono le opere riferite. Ad esempio, il campo 767 specifica informazioni relative alle diverse traduzioni esistenti dell'opera.

**80X–830** Elementi aggiuntivi di una serie. Le serie, specificate nella categoria 4XX, richiedono in alcuni casi degli elementi aggiuntivi, che sono gestiti nella categoria 8XX.

**841–88X** Questi campi contengono informazione specifica per holdings.

**9XX** Lasciato per uso locale

Alcune di queste categorie sono utilizzate come punti d'accesso ai record della biblioteca. Questi, tradizionalmente corrispondono agli elementi che nelle biblioteche tradizionali venivano utilizzati per creare delle schede dell'opera, e che permettevano di effettuare le ricerche. I punti d'accesso corrispondono ai campi 1XX, 4XX, 6XX, 7XX, 8XX.

Per questi campi è importante che i valori inseriti siano validati, in modo da permettere una ricerca più agevole e precisa. Infatti, se i catalogatori fossero completamente liberi nell'inserimento di valori associati a questi campi, le ricerche potrebbero risultare difficoltose: ad esempio, se un catalogatore utilizza come nome dell'autore dell'opera Gabriele D'Annunzio e l'utente cerca le opere di G. D'Annunzio, la ricerca potrebbe non dare alcun risultato. Pertanto risulta necessario che i valori associati a questi campi siano inseriti in un modo ben codificato – che viene poi utilizzato anche in fase di formulazione delle richieste. Lo strumento che viene utilizzato a questo scopo è l'authority control, che consiste in una serie di regole e di termini specifici. Per esempio, per i nomi viene utilizzato un Name Authority file, che contiene la lista di tutti i nomi che possono essere utilizzati nei campi del record MARC<sup>9</sup>. Analogamente esistono liste di nomi geografici, di soggetti, ecc.<sup>10</sup> In generale, l'authority control può differire per le diverse organizzazioni; basti infatti pensare alle differenze che esistono tra le modalità di indicare i nomi nelle diverse lingue.

---

<sup>9</sup>Il Name Authority file della Library of Congress si può accedere all'indirizzo <http://authorities.loc.gov/help/name-auth.htm>

<sup>10</sup>I diversi authority file della Library of Congress si trovano all'indirizzo <http://authorities.loc.gov/>

Come abbiamo visto, ogni tag è individuato da tre cifre, di cui la prima indica la tipologia del campo descritto dal tag. Le altre due cifre vengono utilizzate per individuare la categoria del valore corrispondente al tag, per cui tag di tipologie diverse ma con valori appartenenti alla stessa categoria, hanno le due ultime cifre uguali. Supponiamo, ad esempio di avere due tag, il primo indica un *soggetto* (quindi con il tag è 6XX) ed il secondo una *parte principale* (quindi il tag è 1XX); se i valori di entrambi i tag sono nomi di persona, i due tag saranno, rispettivamente 600 e 100; se invece sono nomi geografici, i tag saranno 651 e 151. In generale, possiamo individuare i seguenti tipi di valori dei diversi campi, con i corrispondenti tag:

- X00 Nomi di persona
- X10 Nomi di aziende
- X11 Nomi di meeting
- X30 Titoli
- X40 Titoli bibliografici
- X50 Argomenti
- X51 Nomi geografici

per cui se combiniamo la codifica delle tipologie dei tag con quella delle categorie dei nomi possiamo ottenere il codice completo del tag.

#### 4.2.5 Descrizione di alcuni campi del record MARC

In questa sezione vengono descritti alcuni tra i tanti campi che possono costituire un record MARC<sup>11</sup>.

**001 - Control number.** Numero di controllo assegnato dall'organizzazione che crea, usa o distribuisce il record.

**005 - Data dell'ultima transazione.** Il campo è costituito di 16 caratteri che specificano la data ed l'ora (nel formato ISO 8601) dell'ultima modifica del record.

Ad esempio 19970714083147.0 corrisponde al 14 Luglio 1997 ore 8:31:47.

---

<sup>11</sup>Una lista più dettagliata dei campi MARC è data in MARC21 Concise Format for Bibliographic Data (<http://www.loc.gov/marc/bibliographic/ecbdhome.html>)

**007 - Descrizione fisica generale.** Il campo descrive le caratteristiche fisiche dell'oggetto. La codifica, basata su un numero variabile di caratteri alfabetici, dipende della categoria dell'oggetto, specificata dal primo carattere. Alcuni esempi di categorie sono

- a Map
- c Electronic resource
- m Motion picture
- r Remote sensing image
- s Sound recording
- t Text
- v Video recording

Ad esempio, per la categoria *sound recording*, il campo è composto di 14 caratteri:

- 00 Categoria del materiale (**s** per sound recording)
- 01 Indica la classe specifica di sound recording (ad esempio, *d* corrisponde a *sound disc*, **s** a sound cassette)
- 02 Indefinito
- 03 Velocità della registrazione. Ad esempio, **b** corrisponde a *33 giri*.
- 04 Configurazione dei canali di playback per la registrazione audio. Ad esempio **s** corrisponde a *stereofonia*, **u** ad *unknown*.
- 05 Dimensione delle tracce sul disco. Esempi: **m** corrisponde a *microgroove*.
- 06 Dimensioni del disco o della cassetta o del cilindro. Esempi: **c** per *7 pollici*, **e** per *12 pollici*
- 07 Dimensione del nastro. Esempi: **m** per *1/4 pollice*, **n** se non applicabile.
- 08 Numero di tracce sul nastro. Esempi: **a** per *1 traccia*, **n** se non applicabile.
- 09 Indica il tipo di disco, cilindro o nastro. Esempi: **a** per *master tape*, **m** per *produzione in serie*
- 10 Indica il tipo di materiale utilizzato per la registrazione. Esempi: **c** per *nastro acetato*, **p** per *plastica*
- 11 Tipo di taglio utilizzato per incidere il disco. Ad esempio **1** indica un *taglio laterale*

**12** Caratteristiche delle apparecchiature per il playback. Esempi: **c** per *Dolby B*, **u** se non conosciuto

**13** Indica come il suono è stato acquisito e memorizzato. Ad esempio **e** indica il suono è stato memorizzato in formato analogico.

Sulla base di queste descrizioni si può interpretare il significato del campo 007 del record riportato in figura 4.2.

**245 - Titolo dell'opera.** Questo campo ha due indicatori

**Primo indicatore** Può assumere il valore 0 se non esiste un altro campo titolo per la stessa opera, oppure 1 in caso contrario.

**Secondo indicatore** Numero di caratteri che devono essere trascurati quando si crea un indice per il titolo. Ad esempio, nel caso dell'esempio in figura 4.2, questo indicatore è uguale a 4, per cui nel titolo *The Beatles and Frank ...* bisogna trascurare i primi quattro caratteri.

Alcuni sottocampi permessi sono i seguenti:

**a** Titolo

**c** Statement di responsabilità

**f** Periodo di tempo durante il quale il contenuto è stato creato

**h** Medium

**n** Numero di parti o sezioni

**s** Versione

**300 - Descrizione fisica.** Fornisce una descrizione fisica dell'opera, come ad esempio le dimensioni e il tipo di materiale di cui è composta. Gli indicatori sono indefiniti (quindi il loro valore è #). Alcuni sottocampi permessi sono:

**a** Estensione (ad esempio il numero di pagine, di volumi, di cassette, di dischi)

**b** Altri dettagli fisici, quali velocità di esecuzione, numero di canali audio, ecc.

**c** Dimensioni, espresse in centimetri, millimetri o pollici.

### 4.3 Dublin Core

Dublin Core è un modello di metadati sviluppato per rappresentare risorse in rete – documenti Web in particolare – utilizzando un insieme di elementi semplice e con un significato universalmente accettato. L’iniziativa è nata durante il secondo congresso internazionale sul World Wide Web che si è tenuto a Chicago nell’Ottobre 1994. In quell’occasione è stata evidenziata la difficoltà di ricercare risorse sul web. Questo problema appare chiaro a chiunque abbia tentato di effettuare delle ricerche sul web utilizzando uno qualunque dei motori di ricerca esistenti: spesso vengono trovati molti documenti che non sono rilevanti per l’utente, mentre quelli rilevanti o non vengono trovati oppure si trovano dopo un gran numero di documenti non significativi. Inoltre, non è facile determinare quanti documenti, potenzialmente rilevanti per l’utente, non vengono individuati perchè non indicizzati dal motore di ricerca. Questo problema, difficilmente risolvibile per tutti i documenti presenti nel web, può essere affrontato nel caso delle Biblioteche Digitali, che pur presentando molti degli inconvenienti che si incontrano sul web per quanto riguarda la ricerca dei documenti, offrono un ambiente controllato, nel quale è possibile associare ad ogni risorsa un insieme di metadati che siano semanticamente significativi.

Le discussioni iniziali che si sono tenute nel corso della conferenza hanno portato all’organizzazione di un Workshop per definire le modalità per la descrizione delle risorse web, in particolare della loro semantica. Durante il workshop, tenutosi a Dublin nell’Ohio nel Marzo del 1995, si è giunti alla conclusione che un ben definito nucleo di elementi semantici può essere appropriato per descrivere delle risorse web, al fine di facilitarne la ricerca. Questo nucleo di risorse è stato chiamato ”Dublin Core metadata set”, dal nome della città che ospitava il convegno. Successivamente, altri workshop simili sono stati organizzati periodicamente, per dibattere delle evoluzioni e dei miglioramenti da apportare al modello, tuttora in evoluzione. Informazioni dettagliate su Dublin Core sono pubblicate sul web a cura della Dublin Core Metadata Initiative (DCMI)<sup>12</sup>.

Il modello Dublin Core è nato per descrivere documenti web, ma è abbastanza generale per poter rappresentare anche il contenuto di altre risorse, quali documenti testuali, suoni, immagini tridimensionali, video, fotografie, film. Attualmente, Dublin Core viene utilizzato da diverse organizzazioni sia private che pubbliche sia come formato di rappresentazione dei metadati che

---

<sup>12</sup><http://www.dublincore.org>



come formato per lo scambio delle descrizioni delle risorse; questo avviene, per esempio, per la Open Archive Initiative (OAI)<sup>13</sup>, <sup>14</sup>.

Il Dublin Core Metadata Set è un insieme di elementi descrittivi capaci di rappresentare qualsiasi risorsa di informazione accessibile in rete. 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 – una pagina web, un audio, un video etc. – che abbia un indirizzo web.

Lo standard Dublin Core comprende un livello di base (*simple*) ed un livello che fornisce informazione più dettagliata (*qualified*) sulle risorse. Il *Simple Dublin Core* è costituito di 15 elementi, che descriveremo in dettaglio nelle prossime sezioni. Poichè il *Simple Dublin Core* può risultare poco specifico ed alle volte ambiguo, il modello è stato esteso, ottenendo il *Qualified Dublin Core*, che comprende un elemento aggiuntivo – Audience – ed un insieme di qualificatori che specificano la semantica degli elementi di base.

#### 4.3.1 Descrizione generale

Il modello Dublin Core è costituito da un insieme, non strutturato – quindi il modello è piatto ed ogni elemento non può avere delle parti – e limitato di 15 elementi. Il modello, data la sua semplicità e generalità può essere utilizzato per descrivere qualunque tipo di risorsa.

I record di descrizione delle risorse sono facili da creare, anche da parte di personale non specializzato (a differenza di quanto succede per i record catalografici di tipo MARC) e sono di uso generale per la "ricerca" delle risorse di informazione (a differenza delle descrizioni catalografiche che in genere seguono normative specifiche del settore di applicazione: biblioteche, musei, videoteche).

Dublin Core è stato definito seguendo un insieme di linee guida, brevemente illustrate nel seguito:

- **Semplicità.**

La creazione e la manutenzione dei metadati deve essere semplice – cioè deve essere possibile anche ai non esperti creare e gestire i metadati

---

<sup>13</sup><http://www.openarchives.org/>

<sup>14</sup>Lo scopo della OAI è di permettere l'accesso trasparente ad archivi diversi contenenti materiale disponibile sul web; questo viene ottenuto attraverso la pubblicazione e la condivisione dei metadati dei diversi archivi.

Dublin Core. Allo stesso tempo, l'insieme di metadati Dublin Core deve essere permettere una efficace ricerca delle informazioni.

Questo criterio, anche se ovviamente auspicabile, può essere difficilmente soddisfatto nella sua interezza; infatti, avere contemporaneamente un insieme di metadati semplice – quindi limitato – ed efficace per l'indicizzazione e la ricerca è estremamente difficile. In generale, Dublin Core ha privilegiato la semplicità alla potenza dell'indicizzazione. Il meccanismo dei qualifiers ha, come vedremo in seguito, lo scopo di aggiungere informazioni di dettaglio, dove necessario.

- **Utilizzo di una semantica universalmente accettata.**

La possibilità di formulare richieste basate su un insieme di metadati richiede che il significato (semantica) che viene associato ai diversi elementi dell'insieme di metadati sia lo stesso per chi effettua la catalogazione e per chi formula le richieste. Poiché le richieste possono essere formulate da utenti molto diversi – per età, cultura, origini, ecc. – questo implica che la semantica dei vari elementi di Dublin Core deve essere conosciuta ed accettata da tutti gli utilizzatori.

Ancora una volta, questo è un criterio che impone una semplificazione del modello, a scapito della precisione e della possibilità di descrivere in modo preciso e dettagliato le risorse. Infatti, anche se questo criterio permette a tipologie di utenti diverse, ad esempio ad uno studente e ad un ricercatore, di formulare richieste su argomenti simili, di fatto impedisce a chi ha bisogno di esprimere richieste complesse o molto precise di operare efficacemente.

- **Possibilità di utilizzo in ambito internazionale.**

Poiché ci si aspetta che Dublin Core venga utilizzato a livello internazionale, sono state approntate diverse versioni nazionali del modello. La nazionalizzazione di Dublin Core è piuttosto semplice, poiché richiede solo la traduzione di ogni elemento del modello e delle varie descrizioni dell'elemento nelle diverse lingue.

- **Estendibilità**

Come si è detto più volte il set di metadati di Dublin Core è piuttosto limitato ed è utilizzabile principalmente per un uso generale e non specialistico. In ogni caso, il modello Dublin Core prevede la possibilità di estendere il set di elementi per soddisfare le esigenze di particolari comunità di utenti.

Il modello Dublin Core è inoltre conforme ad una serie di principi generali:

- Ogni elemento è opzionale e può essere ripetuto.

La possibilità di avere tutti gli elementi opzionali permette di gestire facilmente l'interoperabilità con altri modelli – infatti sarà sempre possibile convertire un modello di metadati qualunque in Dublin Core, a patto di accettare una perdita di informazioni – e permette di semplificare la verifica della correttezza sintattica dei record Dublin Core.

Comunque, questa caratteristica ha degli svantaggi; in particolare, l'estrema flessibilità lasciata nella creazione dei record di metadati, permette di costruire descrizioni delle risorse poco significative: si pensi ad esempio ad una risorsa descritta solo attraverso il campo *data*, o attraverso il campo *tipo*, senza alcun *titolo* né *autore*.

- Ogni record di metadati Dublin Core descrive una manifestazione della risorsa.

Spesso succede che una risorsa sia correlata ad un'altra, o in modo esplicito o implicitamente. Per esempio, la foto della Torre di Pisa è strettamente correlata alla Torre di Pisa ma i metadati Dublin Core della foto devono essere relativi alla foto stessa, piuttosto che alla Torre di Pisa. Quindi il metadato *autore* si riferisce all'autore della foto e non al costruttore della Torre. Potremo utilizzare un elemento specifico di Dublin Core – *relazione*, che verrà presentato nel seguito – che permette di descrivere le relazioni con altre risorse. In questo modo, l'utente che cerca una foto della Torre di Pisa troverà esattamente la foto che stiamo descrivendo e potrà anche accedere ad informazioni sulla Torre di Pisa, qualora fosse interessato.

- La presenza dei qualificatori può essere ignorata.

Come abbiamo visto Dublin Core permette di associare dei qualificatori ad ogni record di elementi che descrivono una risorsa. I qualificatori sono un'estensione di Dublin Core e potrebbero non essere supportati da alcuni sistemi. Per rendere l'utilizzo di Dublin Core il più generale possibile, si richiede che qualunque qualificatore possa essere ignorato e che l'elemento possa essere utilizzato come se il qualificatore non esistesse. In questo caso la descrizione della risorsa risulta meno precisa, ma è comunque sempre possibile utilizzarla.

### 4.3.2 Dublin Core metadata set

Come accennato nell'introduzione, il Dublin Core Metadata Element Set è costituito da quindici elementi. La descrizione di questi elementi è reperibile nel documento Dublin Core Metadata Element Set, Version 1.1: Reference Description<sup>15</sup>, mentre la traduzione italiana è stata curata dall'ICCU<sup>16</sup>.

Nel seguito diamo una breve descrizione di tutti gli elementi costituenti il Dublin Core Metadata Element Set; per ogni elemento riportiamo il nome, l'identificatore, la definizione ed alcuni esempi. Al termine di questa sezione riportiamo alcuni esempi di uso di Dublin Core per la descrizione di risorse diverse.

#### 1. 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 = "2001: Odissea nello spazio"

**Commento:** Il solo titolo non permette di individuare in modo univoco la risorsa. Nell'ultimo esempio, non siamo in grado di dire se si tratta del libro "2001: Odissea nello spazio" o del film avente lo stesso titolo.

#### 2. Nome: Creatore

**Identificatore:** Creator

**Definizione:** Un'entità che ha la responsabilità principale della produzione del contenuto della risorsa. Il creator non deve necessariamente essere una persona, ma può anche essere un'organizzazione, un servizio software, ecc..

**Regole generali:** Se vi sono diversi creator per la risorsa, questi devono essere riportati separatamente e nel caso di persone, il cognome deve precedere il nome. Nel caso di organizzazioni strutturate, bisogna elencare le varie componenti, a partire da quella principale.

**Esempi:** Creator = "Alighieri, Dante"

Creator = "Internal Revenue Service. Customer Complaints Unit"

---

<sup>15</sup><http://dublincore.org/documents/dces/>

<sup>16</sup><http://www.iccu.sbn.it/metadati.htm>

### 3. Nome: **Soggetto e Parole chiave**

**Identificatore:** Subject

**Definizione:** Il Subject è l'argomento della risorsa. Tipicamente un Subject 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.

**Commento:** Non è possibile avere Subject strutturati, del tipo "-software/biblioteche digitali", ma sarà necessario utilizzare due subject diversi, con un'ovvia perdita di informazione.

**Esempi:** Subject = "Poesia"

Subject = "Sinfonia"

Subject = "Software"

### 4. Nome: **Descrizione**

**Identificatore:** Description

**Definizione:** Una spiegazione del contenuto della risorsa

**Commento:** Una descrizione può includere un riassunto analitico, un indice, un riferimento al contenuto di una rappresentazione grafica o un testo libero del contenuto, ecc. Ovviamente, essendo una descrizione non vincolata da alcun limite, il suo contenuto può essere poco specifico o preciso.

**Regole generali:** È molto utile in fase di ricerca. Normalmente vengono utilizzati testo libero e frasi complete.

**Esempi:** Description = "Questa è un'immagine della Piazza dei Miracoli a Pisa. Si può vedere, sulla destra la Torre di Pisa, al centro il Duomo e sulla sinistra il Battistero".

### 5. Nome: **Editore**

**Identificatore:** Publisher

**Definizione:** Un'entità responsabile della produzione della risorsa.

**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"  
Publisher="Università degli Studi di Pisa"

6. **Nome: Autore di contributo subordinato**

**Identificatore:** Contributor

**Definizione:** Un'entità responsabile della produzione di un contributo relativo 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.

7. **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.

**Esempi:** Date= "11 Maggio 2001"

Date= "11052001"

Date= "20010511"

**Commento:** Normalmente si possono associare diverse date ad una risorsa, relativamente ad aspetti diversi, quali la data di creazione, la data di ultima modifica, ecc. Dublin Core, nella sua forma semplificata non permette di distinguere tra le diverse date. Inoltre, Dublin Core non impone nessun vincolo per la codifica della data – specificando se debba essere espressa ad esempio nella forma gg-mm-aaaa, o aaaa-mm-gg – con conseguente ulteriore ambiguità di interpretazione.

8. **Nome: Tipo di risorsa**

**Identificatore:** Type

**Definizione:** La natura o il genere del contenuto della risorsa

**Commento:** Questo campo 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`.

I valori che possono essere associati all'elemento *type* sono specificati nel modello Dublin Core<sup>17</sup>. La lista ed una descrizione dei valori permessi viene fornita alla fine della lista degli elementi di Dublin Core, in questa sezione.

**Esempi:** `Type="Text"`

`Type="Sound"`

`Type="Image"`

## 9. Nome: Formato

**Identificatore:** `Format`

**Definizione:** La manifestazione fisica o digitale della risorsa.

**Commento:** Normalmente `Format` specifica il tipo di supporto o le dimensioni della risorsa. Il valore del matadato `Format` può essere usato per determinare il software, l'hardware o l'apparato necessari alla visualizzazione o all'elaborazione della risorsa. 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 che definisce i formati dei supporti elettronici. Se vi sono più formati per una stessa risorsa, il campo `Format` deve essere ripetuto.

**Commento:** Esempi dei MIME media types sono *plain text*, *html text*, *message*, *jpeg image*, *audio*, *mpeg video*, ecc. È disponibile una lista completa dei MIME media types<sup>18</sup>.

**Esempi:** `Title="Dublin Core icon"`

`Identifier="http://dublin-core/images/dc2.gif"`

`Type="Image"`

`Format="image/gif"`

`Format="4 kB"`

## 10. Nome: Identificatore della risorsa

**Identificatore:** `Identifier`

---

<sup>17</sup><http://dublincore.org/usage/terms/dcmitype/>

<sup>18</sup><http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>

**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"

#### 11. 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"

#### 12. Nome: **Lingua**

**Identificatore:** Language

**Definizione:** La lingua del contenuto intellettuale della risorsa. Si raccomanda l'uso di valori presi da una lista standard <sup>19</sup>.

**Esempi:** Language="it"  
Language="Primarily English, with some abstracts also in French"

#### 13. 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.

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

---

<sup>19</sup>Network working group. Tag for the identification of Languages.  
<http://www.ietf.org/rfc/rfc3066.txt>



#### 14. **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) 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.

#### 15. **Nome: Gestione dei diritti**

**Identificatore:** Rights

**Definizione:** Informazione sui diritti esercitati sulla risorsa

**Commento:** Normalmente un elemento Rights contiene un'indicazione sulle modalità previste per la 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.

**Esempi:** Rights="Accesso limitato ai membri dell'associazione"

Come detto nello standard si suggerisce di utilizzare per l'elemento *type* solo i valori selezionati da un vocabolario di termini. La lista completa di questi valori è la seguente:

##### 1. **Nome:** Collection

**Definizione:** Una *collection* è una aggregazione di elementi. Specificando che il tipo della risorsa è *collection*, si intende quindi che la risorsa è descritta come un insieme di elementi raggruppati. Le singole parti dell'insieme potranno essere descritte separatamente.

##### 2. **Nome:** Dataset

**Definizione:** Un *dataset* è informazione codificata in una struttura definita (ad esempio una lista di elementi, una tabella, ecc.).

3. **Nome:** Event

**Definizione:** Una risorsa di tipo *event* è non persistente e viene creata sulla base di un avvenimento temporale. Esempi di risorse di tipo *event* sono *exhibition*, oppure *workshop*.

4. **Nome:** Image

**Definizione:** Rientrano in questa categoria tutte le immagini grafiche, le foto, le animazioni, i filmati, i quadri, i disegni, ecc. Le risorse di tipo *image* possono includere sia rappresentazioni in formato elettronico che rappresentazioni fisiche.

5. **Nome:** Interactive Resource

**Definizione:** Una *interactive resource* richiede l'interazione con un utente. Esempi sono *oggetti multimediali interattivi*, *applet*, *form nelle pagine web*.

6. **Nome:** Moving Image

**Definizione:** Una *moving image* è una risorsa che, quando viene visualizzata, dà l'impressione del moto. Esempi sono le *animazioni*, i *filmati*, i *programmi televisivi*. A risorse alle quali viene assegnato il tipo *Moving Image* può anche essere assegnato il tipo *Image*, che ha un significato più generale.

7. **Nome:** Physical Object

**Definizione:** Un *Physical Object* si riferisce ad un oggetto fisico reale, non alla sua rappresentazione digitale. Per esempio, un computer, una statua. La rappresentazione digitale di un *Physical Object* può essere un'immagine, una descrizione testuale, ecc.

8. **Nome:** Service

**Definizione:** Un *Service* è un sistema che fornisce un insieme di funzionalità ad un utente. Esempi sono i *servizi bancari*, un *servizio di prestito di libri*, un *servizio web*.

9. **Nome:** Software

**Definizione:** Il tipo *Software* si utilizza per i programmi per computer – sia sorgenti che compilati – che possono essere disponibili

per l'installazione. Questo tipo non va utilizzato per i programmi che creano un ambiente interattivo, per i quali bisogna utilizzare il tipo *Interactive*, visto in precedenza.

10. **Nome:** Sound

**Definizione:** Le risorse di tipo *Sound* vengono presentate in forma sonora. Esempi di risorse di tipo *Sound* sono un *file musicale*, un *CD audio*, una *registrazione di parlato*.

11. **Nome:** Still Image

**Definizione:** Una *Still Image* è una rappresentazione visuale statica. Esempi sono *disegni*, *quadri*, *disegni grafici*.

Nel caso di immagini di materiale testuale, si raccomanda di usare il tipo *Text*. Istanze del tipo *Still Image* possono essere descritte con il tipo più generale *Image*.

12. **Nome:** Text

**Definizione:** Una risorsa di tipo *Text* è costituita principalmente da parole. Esempi sono *libri*, *articoli*, *riviste*. Si noti che una immagine che rappresenta una pagina di testo (ad esempio una immagine acquisita da scanner) è da considerare di tipo *Text*.

### 4.3.3 Dublin Core con qualifiers

In molti casi, gli elementi di Dublin Core sono poco specifici ed ambigui, per cui è stata evidenziata la necessità di fornire informazioni più dettagliate su alcuni campi. Questa esigenza è stata considerata dal gruppo che ha definito lo standard e che ne cura l'evoluzione, per cui sono stati introdotti dei qualificatori (*qualifiers*) che possono essere associati ad ogni elemento di Dublin Core. Come già ricordato, l'uso dei qualificatori è opzionale, sia in fase di inserimento dei metadati che per quanto riguarda il loro utilizzo, per cui ogni programma che analizza una descrizione Dublin Core deve essere in grado di ignorare i qualifiers.

Vi sono attualmente le seguenti due classi di qualifiers:

**Element refinement.** Forniscono delle informazioni aggiuntive in modo da rendere il significato di un elemento più specifico. L'uso di un *element refinement* per un dato elemento non ne modifica la semantica, ma ne precisa meglio il significato. Ad esempio, possiamo avere dei refinement per il campo *Title*, per specificare titoli alternativi – come ad esempio delle abbreviazioni. Oppure possiamo avere dei refinement

per il campo *Date*, per specificare se si tratta della data di creazione (*Created*), della data di modifica (*Modified*), della data di sottomissione (*Submitted*), ecc. Una lista completa degli *element refinements* è disponibile sul web<sup>20</sup>.

Un utilizzatore, sia esso una persona o un programma, che non ha conoscenza del significato del *refinement*, deve comunque essere in grado di operare correttamente sull'elemento.

**Schemi di codifica** (Encoding scheme). In molti casi l'interpretazione del valore di un elemento può risultare ambigua. Si pensi ad esempio al campo *data*, per il quale è necessario conoscere precisamente il formato per una sua corretta interpretazione: la stringa 04-01-02 risulta ambigua se non si sa se la sequenza corretta è aa-mm-gg oppure gg-mm-aa oppure un'altra ancora.

In altri casi può essere utile specificare un vocabolario dei valori permessi per un dato campo: per esempio, potremmo avere una lista dei nomi consentiti per gli autori, oppure una lista dei nomi dei luoghi permessi. In questo modo l'inserimento dei metadati risulta semplificato ed è possibile specificare meglio le richieste degli utenti durante le interrogazioni.

#### 4.3.4 Dublin Core ed XML

Dublin Core può essere rappresentato utilizzando diversi formati. Sono state sviluppate rappresentazioni sia di "simple DC" che di "qualified DC" in XML<sup>21</sup> ed RDF<sup>22, 23</sup>. Esiste inoltre una mappatura di Dublin Core con UNIMARC<sup>24</sup>.

---

<sup>20</sup><http://www.dublincore.org/documents/usageguide/qualifiers.shtml>

<sup>21</sup><http://dublincore.org/documents/2003/04/02/dc-xml-guidelines/>

<sup>22</sup><http://dublincore.org/documents/2002/07/31/dcmes-xml/>

<sup>23</sup><http://dublincore.org/documents/dcq-rdf-xml/>

<sup>24</sup><http://www.iccu.sbn.it/dubluni.html>