

Convegno
"La biblioteca ibrida: verso un servizio informativo integrato"
Milano, Palazzo delle Stelline, 14 -15 marzo 2002.

Metadati:

parola chiave per l'accesso alla biblioteca ibrida

di Antonella De Robbio¹

"...Ma anche questo è solo un passo intermedio.
Già da molto tempo i bibliotecari hanno capito che
il modo migliore per accedere con facilità alle informazioni
non consiste nel cercarle direttamente,
ma nel restringere la ricerca a insiemi mirati e ristretti di dati:
da qui nascono gli schedari delle biblioteche.
Queste informazioni che riguardano altre informazioni
sono chiamate metadati"

[John Bosak e Tim Bray "XML e la Rete di seconda generazione",
Le Scienze 373/settembre 1999]²

Abstract

La presentazione di questo lavoro sui metadati si sviluppa entro tematiche offerte sotto forma di FAQ (Frequently Asked Questions) entro il quadro definito come "biblioteca ibrida". Metadati, cerniera tra la solida tradizione e il futuro tecnologico, ponti tra il vecchio mondo e il nuovo, punti di snodo tra la biblioteca tradizionale e la biblioteca digitale. I metadati presentati e descritti in questo elaborato sono osservati da più punti di vista, che riflettono i diversi modi con cui sono percepiti e suddivisi in categorie dagli autori che ruotano attorno al mondo dell'informazione.

Non si procederà pertanto a stilare liste di formati o standard di metadati, né tanto meno a descrivere il noto Dublin Core nei suoi 15 elementi o il solido e strutturato MARC.

Si tenterà piuttosto di chiarire confusioni e fraintendimenti terminologici, dando una visione "unica" che tenga conto della loro tipologia strutturale, funzionale e delle relazioni con i domini di appartenenza.

Come? Cosa? Dove? Quanti? Chi? Quando? Perché?...

¹ Biblioteca Seminario Matematico. SBA (Sistema Bibliotecario di Ateneo) Università degli Studi di Padova
mailto:antonella.derobbio@unipd.it

²John Bosak e Tim Bray "XML e la Rete di seconda generazione". In "Le Scienze" 373/settembre 1999. Articolo originale "XML and the Second-Generation Web" In "Scientific American", May 1999
<<http://www.sciam.com/1999/0599issue/0599bosak.html>>

La biblioteca ibrida presenta le sue risorse

1. Come si esprimono i metadati?

La maggior parte dei documenti e delle collezioni di documenti hanno in associazione ai loro contenuti qualcosa che è noto come metadato.

La rappresentazione di forme e contenuti avviene attraverso i metadati, ma se nella biblioteca tradizionalmente intesa possiamo parlare di "descrizione delle risorse" o dei documenti, nella biblioteca digitale dobbiamo prima di tutto parlare di "presentazione delle risorse".

Un'efficace gestione dell'informazione digitale deve necessariamente includere la presentazione delle risorse, la gestione degli accessi basata sulle informazioni sui diritti di proprietà, la conservazione a lungo termine. Queste applicazioni portano ad aumentare il consenso attorno ai metadati, consentendo uno sviluppo efficace e conseguente utilizzo di sistemi che possano collezionare e usare metadati appropriati.

La "presentazione" dei metadati di una risorsa avviene, più o meno, nello stesso modo in cui, in biologia, avviene la presentazione di uno o più antigeni da parte di una cellula o di una proteina. La cellula "presenta" l'antigene "esponendolo" entro un contesto noto, di modo che vettori specifici, di "forma compatibile", possano individuarlo ai fini della cattura, agganciandolo e recuperando la proteina. Proteina e antigene sono strettamente correlati: l'uno espone l'altro e la presentazione di certe informazioni consente l'ingresso da parte di agenti esterni nella proteina stessa. Nello stesso modo funzionano i metadati nel Web, il documento o la risorsa digitale espone il metadato, come fosse un suo antigene. Il metadato, che presenta la risorsa e che contiene le informazioni utili, consente a vettori specifici di essere individuato ai fini della localizzazione della risorsa. Se vi è compatibilità di forme e di linguaggi, avviene l'aggancio ai fini del recupero. Per compatibilità possiamo, procedendo per analogia, intendere interoperabilità. L'aggancio è reso possibile dagli "identificatori" contenuti nei metadati. *"E' importante capire"* sottolineare Juha Hakala³ *"che nell'immediato futuro il modo con il quale identificheremo le risorse in generale e quelle elettroniche in particolare, cambierà in modo sostanziale e questo avrà un impatto considerevole sui nostri sistemi e sullo staff"*. I metadati sono essenziali per l'identificazione delle risorse e il funzionamento dei servizi di risoluzione, ci dice Hakala e gli identificatori sono la chiave nella chiave.

Gli identificatori si dividono in due grandi aree: quelli tradizionali come ISBN; ISSN, SICI, BICI, e quelli emergenti che si riferiscono alle "opere" intese nel senso di FRBR. I primi necessariamente dovranno far posto, nelle loro strutture, alle risorse elettroniche e perciò necessariamente dovranno essere ampliati o rivisti nel loro impianto strutturale. I secondi sono necessari per riunire assieme, sotto un unico identificatore, le differenti manifestazioni in cui un'opera può manifestarsi. La gestione di questi identificatori è demandata ai "resolvers" o "risolutori"⁴. Entro questi "servizi di risoluzione" i risolutori, attraverso gli identificatori contenuti nei metadati, si occupano del collegamento tra le descrizioni della risorsa e la risorsa stessa.

Sistemi globali per la localizzazione di risorse e documenti raggruppano nuclei comunicanti attraverso canali di scambio per il trasporto informativo entro protocolli comuni. Tali nuclei, all'interno dei sistemi di accesso all'informazione, contengono standard di metadati i quali riflettono le peculiarità di comunità di utenti differenti. Per questa ragione i metadati rappresentano la chiave d'accesso a differenti sistemi comunicativi, i quali utilizzano linguaggi diversi.

³ Juha Hakala *"Principi di identificazione: prospettive europee"* Convegno internazionale *Le risorse elettroniche: definizione, selezione e catalogazione*. Roma, 26-28 novembre 2001

<http://w3.uniroma1.it/ssab/er/relazioni/hakala_ita.pdf>

⁴ Un risolutore OpenURL si trova sul sito UKOLN "OpenURL demonstrator"

<<http://www.ukoln.ac.uk/distributed-systems/openurl/>>

Prima ancora, i metadati sono l'anello che congiunge i due mondi, quello della biblioteca tradizionale e le nuove forme nascenti di biblioteca digitale.

L'integrazione tra i due mondi, attuata dallo scambio comunicativo tra metadati, dà vita e origine ad un ambiente che prende il nome di "biblioteca ibrida".

Biblioteca ibrida: servizio di accesso all'informazione

2. Perché usare sistemi di metadati?

Il sistema di metadati è la componente logica di qualsiasi biblioteca, sia essa digitale o tradizionale. A differenza di una biblioteca tradizionale, una biblioteca digitale necessariamente si fonda sulle reali applicazioni di metadati e dei loro formati.

Sia nei cataloghi tradizionali - e qui includo anche gli OPAC Online Public Access Catalogue - sia all'interno di insiemi di metadati di biblioteche digitali - che descrivono risorse elettroniche o oggetti digitali - ciò che riveste importanza strategica è l'organizzazione informativa.

L'organizzazione concettuale dell'informazione negli insiemi di metadati che compongono le biblioteche digitali, contrariamente a quanto si crede e al pari dell'organizzazione dei volumi entro una biblioteca fisica, non è per nulla una questione di secondaria importanza. Caso mai può risultare molto più rigorosa anche rispetto alle analisi condotte fino ad oggi negli ambienti della catalogazione tradizionale.

Con l'incremento dei dati nel Web, ci sono molte iniziative volte ad aggiungere informazioni ai documenti digitali attraverso l'uso di metadati. Nel Web i metadati possono essere usati per molti scopi. Alcuni di questi metadati sono catalografici. Altri, come per esempio quelli per la gestione dei diritti di proprietà intellettuale, o la firma digitale usati ai fini delle autenticazioni, sono utili nelle applicazioni per il commercio elettronico e vengono utilizzati nei domini correlati all'industria e al commercio. Livelli di riservatezza - chi dovrebbe e chi non dovrebbe avere accesso ad un documento - sono sempre gestiti da metadati, in categorie in ordini di contenuto, per esempio per proteggere i minori da letture di un certo tipo riservate ad un pubblico adulto, come avviene in *PICS Platform for Internet Content Selection*⁵.

Sono previsti vocabolari predefiniti per metadati di autorità, in particolare per sistemi di tariffazione su servizi per l'accesso ai contenuti. I metadati sono anche utili per le metadescrizioni di oggetti non testuali, per esempio un insieme di parole chiave che descrivono un'immagine. Queste parole chiave possono successivamente essere usate per ricercare l'immagine usando le tecniche di recupero dell'informazione testuale classica (sulle metadescrizioni), per questa ragione in questo caso i metadati sono importanti anche in relazione alle tecniche di *MultiMedia Information Retrieval* (MMIR)⁶.

Allo stato attuale vi sono molti progetti Web sulle ontologie per differenti domini di applicazioni. L'ingresso di un settore come l'ontologia, territorio di competenza dei filosofi, nelle tecnologie dell'informazione, ha innescato non poche polemiche tra gli informatici, mentre, per il momento, i bibliotecari sembrano rimanere su posizioni di roccaforte.

Tim Berners-Lee afferma che la tecnologia per la rappresentazione della conoscenza è nello stato in cui era l'ipertesto prima dell'avvento del Web: una buona idea che non aveva ancora cambiato il mondo. *"Per realizzare appieno le sue potenzialità, questa deve essere collegata in un unico sistema globale"*⁷

⁵ PICS, impiegato come mezzo per filtrare contenuti, utilizza etichette che prevedono gradi numerici al fine di determinare scale di valori associati a una determinata risorsa [es. livello violenza = 10].

⁶ Roberto Raieli *"MultiMedia Information Retrieval"*. Biblioteche Oggi. Vol. XIX - N. 10 Dic. 2001, pag. 16

⁷ Tim Berners-Lee, James Hendler And Ora Lassila *"The Semantic Web"*. Scientific American.

<<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>>

Le biblioteche digitali riflettono forme e modi, oltre che i contenuti, delle comunità che esse rappresentano, così come avviene nella biblioteca tradizionale, ma con qualche differenza fondamentale.

Metadati e confusione semantica

3. Cosa si dice sui metadati?

Il termine metadato è divenuto un termine alla moda e abusato, ha significati differenti per differenti discipline⁸ e, forse a causa dell'ambiguità di significato, c'è chi non lo ama affatto. Michael Gorman in occasione del recente convegno internazionale di Roma sulle risorse elettroniche si è così espresso in merito ai metadati:

"Poi c'è il problema della catalogazione e dei metadati. Su questi il mio punto di vista è che siano un tentativo sconsiderato di trovare una sorta di terza via tra il deserto dei motori di ricerca e della ricerca a testo libero e la grande architettura del controllo bibliografico che i bibliotecari hanno sviluppato negli ultimi centocinquanta anni. Penso che i metadati siano il prodotto di quelli che, senza alcuna conoscenza della catalogazione, né riguardo per essa, come alchimisti bibliografi cercano la pietra filosofale che ci offrirà una catalogazione efficace senza spese e accesso efficace senza vocabolari controllati. Non esiste nessuna cosa del genere e prima si abbandona meglio è"⁹.

Al lato opposto di affermazioni di principio così categoriche vi sono schiere di esperienze già consolidate, basti vedere in seno UKOLN¹⁰ in Europa o nelle biblioteche digitali di California o nei progetti australiani o in altre parti del mondo ancora, come i metadati siano ormai parte di una realtà quotidiana, tasselli fondamentali di un processo dal quale difficilmente si può tornare indietro. Manzi e Martellini¹¹ ci raccontano il recente convegno internazionale di Roma e così commentano: *"[...] i metadati sembrano, rispetto al lavoro di catalogazione e ai nuovi documenti "risorse elettroniche", la cerniera in grado di coniugare vecchio e nuovo; sembrano ciò che è in grado di opporre una resistenza al carattere mutevole e difficilmente identificabile/individuabile delle risorse elettroniche, con particolare riferimento al loro valore rispetto alla conservazione di questo tipo di documenti."*

Anche Priscilla Caplan scriveva nel 1995:

"I metadati sono nulla di più che dati sui dati: i record dei cataloghi sono metadati: così come un header TEI, o qualsiasi altra forma di descrizione. Potremmo chiamarla catalogazione, ma per alcuni tale termine porta in sé un bagaglio eccessivo, come le Anglo-American Cataloging Rules e USMARC. Così alcuni lo estendono ad una situazione del tipo "Voi chiamatelo granturco, noi lo chiamiamo mais"¹², ma i metadati sono un buon termine neutrale per coprire tutte le basi."¹³

⁸ Rachel Heery, Andy Powell and Michael Day "Metadata". Library & Information Briefings. Issue 75 September 1997.

⁹ Michael Gorman "Electronic resources: which are worth preserving and what is their role in library collections?" Versione in italiano "Le risorse elettroniche: quali meritano di essere conservate e che ruolo svolgono nelle raccolte delle biblioteche?" Convegno internazionale *Le risorse elettroniche: definizione, selezione e catalogazione*. Roma, 26-28 novembre 2001
<http://w3.uniroma1.it/ssab/er/relazioni/gorman_ita.pdf>

¹⁰ Vedi sito UKOLN

<<http://www.ukoln.ac.uk/metadata/biblink/>>

¹¹ Stefania Manzi e Enrico Martellini "Alcune riflessioni in margine al convegno Le risorse elettroniche, Roma 26-28 novembre 2001" La relazione riassume con commenti e riflessioni le significative esperienze internazionali portate al congresso da oltre venti esperti provenienti da tutto il mondo. Su ESB Forum:

<<http://www.burioni.it/forum/manzi-er.htm>>

¹² *Corn e Maize* in inglese sono sinonimi: *corn* è più inglese britannico, *maize* è più americano perché deriva dal

C'è una differente enfasi entro le discipline informatiche, dove il termine si riferisce ai dati i quali descrivono elementi di dati, insiemi di dati o sistemi di gestione di basi di dati.

Per esempio, all'interno di un sistema di gestione di basi di dati, lo schema specifica alcuni dei metadati, cioè il nome delle relazioni, i campi o gli attributi di ciascuna relazione, il dominio di ciascun attributo e così via ... Ancora, espressioni come "modelli di metadati" o "sistemi di metadati" sono costruiti piuttosto per integrare diverse basi di dati.

Si può notare come esistano sovrapposizioni tra la presentazione di risorse e la gestione dell'informazione, ma ci sono marcate differenze nella natura dei dati descritti: l'unità che è stata descritta, in contesto informatico, viene considerata un elemento del dato, mentre nel mondo dell'informazione diviene una risorsa. Nel mondo dell'informazione il metadato può consistere in un insieme omogeneo di elementi di dati con una semantica comune, sintassi convenzionale e regole standardizzate per la formulazione del contenuto degli elementi.

La nuova terminologia ci aiuta. Il termine metadato è utile a riconoscere il significativo cambiamento che viene posto nell'enfasi tra la tradizionale catalogazione e l'attività odierna che tenta di ordinare i materiali digitali originati dall'esplosione del Web.

In particolare il concetto di descrizione di risorsa è differente da quello di presentazione della risorsa. Nella lingua inglese il termine "*discovery*" ha un significato duplice, di scoperta, ma anche di "*presentazione*". Questa duplicità del termine comporta ambiguità di interpretazione quando si passa da una lingua (l'inglese) ad un'altra (italiano).

Nel caso specifico l'ambiguità viene acuita se poniamo il termine "*discovery*" nel Web, correlandolo alle risorse. Viene quasi intuitivo tradurlo come "*scoperta*". In realtà non possiamo parlare di "*scoperta*" di risorse poiché la scoperta solitamente si riferisce a qualcosa che avviene per la prima volta e non alle successive. Nei servizi di accesso all'informazione più che di risorse "*scoperte*", dobbiamo riferirci a risorse "*presentate*". Solo una presentazione adeguata consente l'identificazione e la conseguente localizzazione della risorsa.

Nella lingua inglese il termine "*discovery*" assume, infatti, anche valenza di "*presentazione*", significato di derivazione dal diritto dove, a seguito di una specifica istanza, vengono "*presentati*" i relativi documenti utili. Ad avvalorare questa teoria cito anche Martin Dillon¹⁴ il quale, come vedremo in seguito, differenzia nettamente l'attività di "*presentazione delle risorse*" (*resource discovery*) da quella di "*descrizione delle risorse*" (*resource description*).

Secondo Priscilla Caplan¹⁵, l'uso di termini "nuovi" che non hanno tradizionali connotazioni di catalogazione sono a tutto vantaggio della biblioteconomia, giacché riflettono cambiamenti nella natura stessa della catalogazione acquisita attraverso la tecnologia digitale, cambiamenti che sono stati tipizzati da David Levy come "catalogazioni nell'ordine digitale"¹⁶

Origini e nascita dei metadati

4. Quando sono nati i metadati?

L'origine e primi usi del termine metadato trovano radici nell'ambiente correlato all'informazione digitale, in particolare nel contesto delle basi di dati.

nome azteco della pianta, così come granturco e mais in italiano. Il modo di dire americano "*You call it corn as we call it maize*" sta a significare che si stanno facendo solo giri di parole...

¹³ Priscilla Caplan "*You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects.*"

The Public-Access Computer Systems Review 6, no. 4 (1995).

<<http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>>

¹⁴ Martin Dillon "*Metadata for Web Resources: How Metadata Works on the Web*"

<http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>

¹⁵ Priscilla Caplan, op. cit

¹⁶ David M. Levy "*Cataloging in the Digital Order*", 1995

<<http://www.csd.tamu.edu/DL95/papers/levy/levy.html>>

Una scoperta assai curiosa in cui mi sono imbattuta è stata l'indicazione, come riportato dal W3C¹⁷, che il termine "*metadata*" è un marchio registrato sin dal 1986 di proprietà di *The Metadata Company*¹⁸. La sua invenzione è accreditata, come afferma Dillon, a Jack E. Meyers il quale disse di aver coniato il termine nell'estate del 1969. Il marchio è scritto in lettere tutte maiuscole e senza separazione o trattini (meta data o meta-data).

Ad oggi per metadati si intendono anche i record catalografici più tradizionali, contenuti negli OPAC che descrivono i documenti delle biblioteche, come pure i record in formato proprietario dei database bibliografici.

Risale al marzo 1995 il noto *Metadata Workshop*¹⁹ organizzato da Stuart Weibel di OCLC che diede origine ai metadati per la comunità dei bibliotecari nel modello Dublin Core.

Se scorriamo qualche testo sulle biblioteche digitali, ci accorgiamo subito di come il termine "metadati" non compaia per nulla, per esempio, in testi fondamentali come il Lesk²⁰ del 1997.

Da allora, da un primo nucleo di significato, e grazie all'attività che si è svolta attorno al Dublin Core, i confini attorno al termine sono andati via via allargandosi fino a comprendere qualsiasi genere o modello di informazione descrittiva standardizzata, fino a raggiungere anche ambiti informativi "pregressi" e precisamente anche le risorse non digitali.

Per questa ragione i cataloghi di biblioteca, i servizi di indicizzazione, gli archivi e la documentazione dei musei possono essere tutti interpretati come contenitori di metadati.

Del resto la visione di Stu Weibel era proprio questa: mettere in comunicazione mondi diversi. I vantaggi di questa visione allargata del termine sono principalmente due.

Il primo è quello che tutti i bibliotecari, gli archivisti e gli specialisti della documentazione museale possono utilmente cooperare in modo trasversale entro i propri confini professionali in una visione estesa.

In secondo luogo ciò consente, nel contesto dei beni culturali, di comunicare in modo più efficace tra domini diversi, per tipologie organizzative o funzionali, che hanno però interesse nell'applicazione di metadati: per esempio sviluppatori di software, editori, industrie discografiche, compagnie televisive, produttori di contenuti educativi e quelli che hanno a che fare con informazioni geografiche o geospaziali.

Definizione di metadato

5. Che cosa significa metadato?

Il termine metadato è qualcosa di letteralmente definito come "dato sul dato". Solitamente viene inteso nel significato di "dato strutturato sulle risorse" il quale può essere utilizzato a supporto di un'ampia gamma di operazioni tra cui la descrizione e presentazione delle risorse, la gestione delle risorse informative, e la loro conservazione a lungo termine.

Il metadato è l'informazione sull'organizzazione del dato, sui vari domini del dato, e sulle relazioni tra dato e suoi domini.

Tim Berners-Lee definisce il metadato come "*informazione comprensibile alla macchina su risorse Web o altri oggetti*". Si tratta della definizione data da Berners-Lee nel suo lavoro del 1997 sugli assiomi dell'architettura Web²¹, precedente gli scritti sul Web semantico.

¹⁷ Nelle note di disclaimer del W3C si legge "*METADATA*" is a trademark of the Metadata Company. W3C uses the term "*metadata*" in a descriptive sense, meaning "data about data". W3C is not in any way affiliated with the Metadata Company."

¹⁸ <www.metadata.com>

¹⁹ Metadata Workshop: the Essential Elements of Network Object Description
<<http://dublincore.org/workshops/dc1/>>

²⁰ Michael Lesk "*Practical Digital Libraries: Books, Bytes and Bucks*". Morgan Kaufmann, July 1997.

²¹ Tim Berners-Lee "*Metadata Architecture*". January 1997

La frase "*machine understandable*", che egli considera come punto "chiave", si riferisce a quell'informazione che agenti intelligenti possono usare per renderci la via più facile, nel rispetto di leggi e di principi, assicurando procedure di lavoro più veloci e precise.

Secondo la visione di Berners-Lee l'assioma è questo: **il metadato è un dato.**

Il metadato può essere immagazzinato come dato entro una risorsa.

Allo stesso tempo una risorsa può contenere informazione su sé stessa o su un'altra risorsa.

Un metadato su un documento può trovarsi dentro il documento o dentro un documento separato o, ancora esso può essere trasferito in accompagnamento al documento che rappresenta.

La seconda parte dell'assioma di Berners-Lee dice che: **un metadato può descrivere un metadato.** Questo perché lo stesso metadato può avere attributi tali come i diritti di proprietà o la data di scadenza: in questo caso siamo di fronte ad un meta-metadato.

Metadati interni e metadati esterni

6. Dove sono collocati i metadati?

Semplicisticamente i metadati possono presentarsi come

1. Interni al documento
2. Esterni al documento

Questa separatezza tra metadati esisteva anche prima dell'era digitale, tra la vecchia schedina bibliografica, esterna al volume e la familiare CIP *Cataloging-In-Publication* parte integrate del testo a stampa.

In effetti esiste grande confusione, a livello generale, in relazione a queste due grandi aree.

Come precedentemente anticipato, Dillon²² differenzia i metadati in due grosse aree: nella prima possiamo collocare i metadati come *tag*, nella seconda i metadati come descrittori.

Nella prima area abbiamo metadati che *presentano* la risorsa, nel secondo metadati che *descrivono* la risorsa. La prima area, la più recente, si colloca nella nascita del Web e definisce una forma di metadati tipizzata in base alle "etichette" o *tag* che caratterizzano il contenuto dei campi. Siamo di fronte a *metadati per il Web* i quali, come vedremo in seguito, si collocano nella prima fascia della tabella di sotto riportata e rielaborata sullo schema di Lorcan Dempsey²³, quella dei metadati non strutturati a formato semplice.

In questo caso i metadati sono incarnati all'interno dello stesso documento, all'interno del linguaggio di marcatura che permette la presentazione dell'oggetto in formato digitale; sono così un tutt'uno con la risorsa che essi descrivono. Tali descrizioni interne possono perciò essere automaticamente riconosciute da motori di ricerca, che leggono all'interno i codici sorgenti con cui sono scritti i documenti, individuando la loro localizzazione.

I metadati interni possono essere estratti automaticamente da opportuni sistemi di recupero e indicizzazione, ai fini della creazione di biblioteche digitali basate su un modello distribuito.

Nella seconda area Dillon colloca gli standard di metadati per la descrizione delle risorse dove possono esistere sia metadati interni sia esterni al documento.

<<http://www.w3.org/DesignIssues/Metadata.html>>

²² Martin Dillon, op. cit.

²³ Rachel Heery with contributions from Robina Clayphan, Michael Day, Lorcan Dempsey, and David Martin. Metadata formats. Work Package 1 of Telematics for Libraries project BIBLINK (LB 4034)

5. Typology of Metadata

<<http://hosted.ukoln.ac.uk/biblink/wp1/d1.1/doc0005.htm>>

Questo perché ci può essere una descrizione di una risorsa, effettuata attraverso uno standard descrittivo, che può convivere entro la risorsa, dentro la parte HEAD del documento.

Lo standard Dublin Core che descrive il documento con metadati semplici, ma strutturati, abita in questa area. Lo stesso Dublin Core può essere però creato automaticamente in modo del tutto indipendente dal documento che rappresenta, come nel caso per esempio di CORC²⁴ di OCLC. CORC è uno strumento per creare descrizioni bibliografiche di materiali di formato diverso, organizzarle e navigare attraverso esse. In CORC, base di partenza per la prossima generazione di servizi catalografici di OCLC, convivono due formati di metadati diversi a livello strutturale, ossia il Dublin Core e il MARC, ma altri formati saranno implementati in futuro. La creazione di metadati avviene automaticamente e altrettanto automaticamente avviene la conversione tra un formato ad un altro.

La divisione in due grosse aree tracciata da Dillon, differenzia ulteriormente i metadati entro la sezione *header*, in base alla considerazione che la meta-informazione ha due funzioni principali:

1. Fornire un mezzo per la presentazione dell'insieme di dati esistenti, comprensivo delle informazioni sulle condizioni di accessibilità della risorsa
2. Documentare il contenuto, la qualità, e la configurazione dell'insieme dei dati, indicando le sue forme di utilizzo.

Popolazioni di metadati e infodiversità

7. Quanti sono i formati di metadati?

Un'indagine condotta nel 1996 entro il progetto europeo DESIRE²⁵ identificava e descriveva oltre venti "popolazioni" diverse per formati in uso o in corso di sviluppo.

La prima cosa che va sottolineata quando si parla di formati metadati è che sono diversi per natura tipologica, per implementazione, per ambiti di provenienza e per numerosi altri fattori, ma soprattutto per il fatto che essi descrivono oggetti di natura e specie differente.

Il termine "*biodiversità*" in biologia indica i diversi livelli di ricchezza della vita sul nostro pianeta e, in tal senso, è abitudine distinguere tre livelli principali: la diversità intraspecifica genetica, la diversità a livello di specie e la diversità a livello di comunità.

Così avviene anche per i metadati. A tal proposito torna utile sottolineare come molti autori assegnano al termine "*metadiversità*" significato di "*metadato in un contesto di biodiversità che può relazionarsi con un'ampia gamma di tipi di informazione che includono dati bibliografici, dati su modellizzazioni di oggetti di musei, tassonomie sviluppate da biologi, prodotti della ricerca e indagini geospaziali*"²⁶.

Anche Clifford Lynch²⁷ utilizza il termine "*metadiversità*"²⁸ indagando minuziosamente sui concetti di biblioteca digitale, interoperabilità, infrastruttura, sebbene personalmente ritengo

²⁴ CORC può essere considerato la naturale estensione di InterCat e NetFirst, sebbene esso sia uno strumento molto più versatile in quanto integra il MARC con Dublin Core e consente l'esportabilità di record in HTML e XML all'interno di altri sistemi. I record di CORC provengono quindi da InterCat e NetFirst per la fase iniziale ed ora è alimentato dalla registrazioni delle biblioteche che hanno aderito all'iniziativa.

²⁵ DESIRE Project
<<http://www.desire.org/>>

²⁶ Michael Day "*Metadata and biodiversity information: a report from a US symposium on "Metadiversity"*" Reports from an interactive symposium held at Natural Bridge, Virginia on 9-12 November 1998. This article appears in the Web, and not the print, version of Ariadne.
<<http://www.ariadne.ac.uk/issue18/metadiversity/>>

²⁷ Clifford Lynch "*Building digital libraries for metadiversity*"

²⁸ Richard T. Kaser and Victoria Cox Kaser, (eds.) "*Metadiversity: responding to the grand challenge for biodiversity information management through metadata*". Philadelphia, Penn.: National Federation of

che "metadiversità" non sia rappresentativo della diversità tra metadati, a causa del prefisso "meta" che non necessariamente si riferisce ai metadati. "Biodiversity" è una contrazione di "biological diversity", ma nel caso di "metadata diversity" appare poco convincente coniare il termine di "metadiversity" abbinandolo ai metadati che sono una parola composta. Considerando che il concetto di "metadiversity" è sempre associato al concetto di "information", sarei dell'idea di individuare una forma più appropriata nel termine "infodiversity", o "infodiversità" laddove per "info" si intende sia l'informazione "metadato" sia il "dato".

Dalla letteratura internazionale si evince che esistono metodi diversi di approccio ad una categorizzazione dei metadati. In analogia a quanto avviene in merito al concetto di biodiversità anche per i metadati si riscontrano sostanzialmente tre livelli di distinzione:

1. **tipologico strutturale**, o intraspecifico al record
2. **funzionale**, correlato alla funzione stessa del metadato in relazione all'ambiente
3. **di comunità**, in una suddivisione in domini sulla base delle specifiche attività di comunità di utenti.

In questo lavoro voglio quindi tentare di dare una visione "univoca" di metadato, tenendo conto delle sue tre dimensioni, strutturale, funzionale e correlata al suo dominio, interagenti su tre piani distinti, trasversali l'uno all'altro.

Suddivisione dei metadati per tipologia strutturale

8. Come sono strutturati i metadati?

In merito all'analisi sull'esistenza di differenti formati di metadati, il progetto DESIRE effettua una distinzione in tre fasce sulla base delle caratteristiche dei record, effettuata su oltre venti formati di metadati differenti.

Sulla base della suddivisione primaria tracciata da Dillon nelle due grosse aree, qui di seguito propongo una rielaborazione della tabella di Lorcan Dempsey and Rachel Heery²⁹ -basata sulle caratteristiche dei record- da me integrata con i contenuti delle due figure tratte dal lavoro di Michael Day³⁰, sulle tipologie di metadati di presentazione delle risorse e delle caratteristiche associate ai formati di metadati.

La tabella così rielaborata riassume tre approcci differenti (Dillon-Dempsey/Heery-Day) in una visione d'insieme che tiene conto di più punti di vista:

Suddivisione Fasce	Caratteristiche dei record	Caratteristiche associate ai formati	Formati dei record
Area dei metadati come tag = metatag			
Prima Fascia	Formati semplici e non strutturati	Localizzazione Generazione	Lycos Altavista

Abstracting & Information Services, 1999, pp. 131-140
ISBN: 0-942308-51-4.

²⁹ Lorcan Dempsey and Rachel Heery "Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats" nell'ambito del Progetto DESIRE del 19 marzo 1997

³⁰ Michael Day "The metadata challenge for libraries: a view from Europe". 1999 [preprint]
<<http://www.pa.utulsa.edu/nfais/metadiversity/mday.html>>

Indicizzazione a testo pieno	Formati proprietari Indicizzazione su tutto il testo pieno	effettuata dai robot	Yahoo
Area dei metadati come descrittori			
Seconda Fascia	Formati strutturati Standard emergenti Strutture dei campi	Selezione Input manuale o automatico Attributi/coppie di valori	Dublin Core IAFA templates ROADS templates RFC 1807 SOIF LDIF NetFirst
Formati generici a struttura semplice			
Terza Fascia	Formati ricchi Standard internazionali Etichettature di campo elaborate	Valutazione e analisi Creazione manuale Sottocampi Standard Internazionali	ICPSR CIMI EAD TEI headers MARC
Strutture più complesse, domini specifici Parte di un'architettura semantica ampia			

Passiamo ora ad una lettura più attenta della tabella sopra riportata, rimanendo sempre nell'ambito delle caratteristiche dei metadati a livello strutturale.

Prima fascia: i metadati creati dai motori di ricerca

In questa fascia rientrano metadati di formato proprietario relativamente non strutturati, in genere estratti automaticamente dalle risorse e indicizzati per la ricerca. Il tipico ambiente che ospita questo tipo di strutture è il Web entro cui gli attuali motori di ricerca creano formati di metadati semplici attraverso un'indicizzazione su tutto il testo pieno della risorsa. Termini e parole chiave vengono assegnate in modo assai generico. In questo caso i dati hanno poca semantica esplicita e non permettono la ricerca per campo.

Per queste ragioni, i motori sono limitati per le funzionalità relative alle attività di scoperta, funzione tipica dell'ambiente di ricerca accademico-scientifico.

Questi problemi sono ben noti e quindi vi è movimento a due vie: da una parte i servizi che gestiscono i motori di ricerca stanno cercando di migliorare i metadati su cui lavorano, dall'altra parte le comunità specializzate stanno mettendo a punto motori dedicati con funzionalità peculiari alla comunità stessa.

Seconda fascia: metadati generici a struttura semplice

Si tratta di metadati strutturati in campi definiti, i quali permettono una ricerca per campi.

I dati sono strutturati in modo semplice tanto che possono essere creati da utenti non specialisti poiché non richiedono significative conoscenze specifiche della disciplina. e generalmente, ma non necessariamente, sono creati manualmente.

Poiché spesso si riferiscono a standard emergenti questi metadati possono comprendere una varietà di attributi, descrittivi o di altro tipo, soprattutto se si appoggiano a modelli (template) per il trattamento di risorse di genere differente, come per esempio nel caso di IAFA template³¹, dove sono definiti quattordici tipi di risorse in modelli distinti che prevedono differenti attributi³². Il loro limite maggiore è che si tratta di metadati poco dinamici, perché le

³¹ IAFA Templates in use as Internet Metadata
<<http://www.w3.org/Conferences/WWW4/Papers/52/>>

³² I quattordici tipi di template IAFA sono i seguenti: Siteinfo, Larchive, Mirror, User, Organization, Service,

descrizioni tendono a riferirsi a oggetti discreti e non catturano relazioni multiple fra gli oggetti, sebbene siano comunque flessibili se inseriti in una struttura relazionale che consenta una relazione tra metadati descrittivi con profili analoghi.

In ragione della loro "adattabilità" e non specificità questo tipo di metadati rappresenta anche un nuovo approccio d'elezione adottato per incrementare i record di grossi cataloghi. E' il caso per esempio del funzionamento del già citato servizio CORC di OCLC dove il metadato Dublin Core creato automaticamente può essere corretto e manipolato in momenti successivi. Il loro vantaggio, che è anche la loro forza, è che i dati contengono una descrizione dettagliata a sufficienza per consentire ad un utente di valutare la potenziale utilità o interesse di una risorsa senza doverla acquisire o connettersi ad essa. Per questa ragione sono gli standard preferiti dai servizi di accesso all'informazione. Un esempio di questi servizi sono i *Subject-Based Information Gateway* (SBIG) che talvolta utilizzano un proprio formato interno, ma altre volte, come in numerosi programmi e-Lib, usano i template IAFA per le descrizioni delle risorse. A fini di una maggiore chiarezza, forse non è scontato affermare che un template è il modello o struttura entro cui avviene la descrizione di una risorsa: il prodotto di tale descrizione è il metadato descrittivo che si differenzia in base alla tipologia del template, che corrisponde alla tipologia della risorsa. Tali modelli solitamente sono flessibili e consentono aperture verso descrizioni di svariate tipologie di risorse. Su questi modelli si basano le architetture, solitamente di tipo relazionale, su cui si costruiscono i portali disciplinari o SBIG. Uno strumento per la creazione di portali, simili all'architettura Yahoo!, è *ROADS Resource Organisation And Discovery in Subject-based services*³³ che mette a disposizione strumenti per l'identificazione, la valutazione, la descrizione e l'organizzazione, per settori disciplinari o per ambiti geografici, di RER³⁴. *ROADS* è scritto in *Perl* con software interamente libero. I programmi sviluppati al suo interno sono messi a disposizione liberamente per usi e implementazioni pubblici. Per la ricerca utilizza il protocollo *WHOIS++*, mentre per la catalogazione implementa i modelli IAFA, il cui schema è già stato mappato col Dublin Core.

A differenza dell'ambiente Web, spesso questi servizi sono orientati a una certa selettività riguardo ai materiali che descrivono e possono avere criteri più o meno espliciti e dichiarati in merito alla selezione. Per queste ragioni la loro realizzazione può presentare dei costi tali da motivare l'interesse a utilizzare descrizioni prodotte dagli autori o dagli editori, nonché tecniche di estrazione automatica.

In quest'ottica è sorto e ha acquisito ampio successo entro numerose comunità il Dublin Core, concepito quale formato strutturato semplice ad utilizzo degli stessi creatori di risorse (autori o editori), mentre sull'altro versante si collocano i servizi per l'estrazione e raccolta automatica di metadati. La filosofia OAI Open Archive Initiative³⁵ rientra in questa tipologia di servizi.

La terza fascia: metadati ricchi a struttura complessa

Comprende formati descrittivi ricchi, strutture più complesse utilizzate per la selezione delle risorse e presentazione adeguata agli scopi. La struttura di questi metadati comprende anche sottocampi con etichettature elaborate. Uno dei più importanti formati di metadati che rientra in questa fascia è il noto MARC il quale è il formato più usato per i record bibliografici poiché è fortemente strutturato essendo dotato di numerosi campi e differenti attributi.

Una delle caratteristiche principali è quella che la creazione è delegata a terze figure, bibliotecari, documentalisti o figure preposte alla selezione dei documenti che devono essere descritti, solitamente dentro un database.

Document, Image, Software, Mailarchive, Usenet, Sound, Video, Faq

³³ <<http://www.ilrt.bris.ac.uk/roads/>>

³⁴ Risorse elettroniche remote

³⁵ <www.openarchive.org>

La scelta delle risorse non è quindi casuale e implica comunque una creazione dei metadati quasi esclusivamente manuale. Gli standard utilizzati per questi metadati sono solitamente internazionali. Oltre che dettagliati nella loro presentazione questi metadati sono sufficientemente espressivi da catturare una varietà di relazioni a livelli differenti.

Solitamente hanno anche un ruolo nella documentazione di oggetti o, molto spesso, di collezioni di oggetti. In tal caso assolvono alla funzione specifica che li definisce metadati di conservazione entro una classificazione per tipologia funzionale.

I metadati di questa fascia sono generalmente associati con attività di studio e ricerca, richiedono conoscenze specialistiche per la loro creazione e mantenimento e alimentano requisiti propri di domini specialistici. In tal caso in virtù delle attività ai quali si riferiscono si riconducono a domini specifici, come vedremo in seguito nella suddivisione dei metadati per domini.

Suddivisione dei metadati per tipologia funzionale

9. Qual è la funzione dei metadati?

I metadati mutano in relazione al mutare degli oggetti a cui essi si riferiscono.

Come notato precedentemente, i metadati non sono solo usati per la descrizione delle risorse o per scopi di presentazione. Non esistono solo metadati descrittivi, che descrivono o catalogano un oggetto, ma possono essere necessarie informazioni che indicano il contesto, la gestione, i processi, la conservazione e l'uso della risorsa che è stata descritta. Per esempio i metadati possono essere usati per scopi amministrativi o per la gestione delle risorse o per registrare informazioni sulla loro acquisizione e localizzazione o quelle sull'acquisizione dei diritti di proprietà intellettuale e aiutare a gestire gli accessi per profili di utenza.

Vi sono altri metadati che hanno una natura tecnica in quanto documentano come le risorse si correlano ad un particolare ambiente software e hardware.

I metadati tecnici sono quelli per esempio che consentono la registrazione dei parametri di digitalizzazione. In questo caso potrebbero correlarsi con i metadati di conservazione.

La creazione e manutenzione di metadati è anche vista come fattore importante nella gestione della conservazione a lungo termine delle risorse digitali e sono utili nella conservazione del contesto e autenticità delle risorse. Esempi tipici di questi formati "ricchi" di metadati sono gli standard per risorse audio-visuali³⁶.

La specificità nel determinare le categorie di metadati è dovuta a vari fattori.

Innanzitutto la dinamicità che certi metadati hanno assunto in relazione ad altri è stato un fattore che ha visto l'emergere di nuovi gruppi di metadati con funzioni differenti ma strettamente correlati ai precedenti.

La tabella di sotto riporta i diversi tipi di metadati differenziandoli sulla base delle loro funzioni, in cinque distinte categorie³⁷

Tipologia	Definizione
------------------	--------------------

³⁶ per esempio MPEG-7 *Multimedia Content Description Interface*, o per immagini digitali.

³⁷ Introduction to Metadata: Pathways to Digital Information, a cura di Murtha Baca. Getty Standards Program
<<http://www.getty.edu/>>

Amministrativi	Metadati utilizzati nella gestione e nell'amministrazione delle risorse informative <ul style="list-style-type: none"> <input type="checkbox"/> Informazioni sull'acquisizione <input type="checkbox"/> Tracciato storico dei diritti di proprietà intellettuale, cessione e passaggi ai fini della riproduzione <input type="checkbox"/> Documentazione dei requisiti di accesso legale <input type="checkbox"/> Informazioni sulla reperibilità <input type="checkbox"/> Criteri di selezione per la digitalizzazione [= formato, set di caratteri] <input type="checkbox"/> Controllo della versione e distinguibilità fra oggetti informativi simili <input type="checkbox"/> Tracce di controllo create da sistemi di gestione di metadati [<i>recordkeeping</i>]
Descrittivi	Metadati utilizzati per descrivere o identificare risorse informative <ul style="list-style-type: none"> <input type="checkbox"/> RegISTRAZIONI catalografiche <input type="checkbox"/> Indicazioni di aiuto per il reperimento <input type="checkbox"/> Indicizzazione su database specialistici <input type="checkbox"/> Connessioni fra risorse tramite link Web <input type="checkbox"/> Annotazioni di utenti <input type="checkbox"/> Metadati per sistemi di gestione [<i>recordkeeping</i>] generati dai programmi di creazione delle registrazioni
Sulla conservazione	Metadati riferiti alla gestione della conservazione delle risorse informative <ul style="list-style-type: none"> <input type="checkbox"/> Documentazione della condizione fisica delle risorse <input type="checkbox"/> Documentazione delle azioni intraprese per conservare le versioni fisiche e digitali delle risorse, per esempio ripristino [<i>refreshing</i>] e migrazione dei dati
Tecnici	Metadati riferiti al funzionamento di un sistema e al comportamento dei metadati <ul style="list-style-type: none"> <input type="checkbox"/> Documentazione sull'hardware e il software <input type="checkbox"/> Informazioni sulla digitalizzazione, per esempio formati, rapporti di compressione, procedure di graduazione [<i>scaling routines</i>] <input type="checkbox"/> Tracciato storico dei tempi di risposta di sistema <input type="checkbox"/> Dati di autenticazione e sicurezza, per esempio chiavi crittografiche, password
Di utilizzo	Metadati riferiti al livello e al tipo di utilizzo delle risorse informative <ul style="list-style-type: none"> <input type="checkbox"/> RegISTRAZIONI di visualizzazione [<i>exhibit records</i>] <input type="checkbox"/> Tracciato storico dell'uso e gestione dei profili utenti <input type="checkbox"/> Informazioni sulla riutilizzo del contenuto e sull'esistenza di una pluralità di versioni [<i>multi-versioning information</i>]

Suddivisione dei metadati per domini di appartenenza

10. Quali sono i domini di appartenenza dei metadati?

Seguendo la classificazione in domini secondo l'Osservatorio dei metadati di SCHEMAS³⁸ sono state suddivise in otto categorie di attività ventuno iniziative di standardizzazione.

Un importante obiettivo del progetto SCHEMAS è fornire informazione e rilevante materiale che funga da guida per il disegno e registrazione di schemi di metadati.

Qui di seguito riporto la tabella che classifica i metadati per domini di appartenenza e ambiti di attività in cui vengono prodotti o utilizzati:

Categoria di dominio	Attività di dominio	Categoria di dominio	Attività di dominio
----------------------	---------------------	----------------------	---------------------

³⁸ <http://www.schemas-forum.org/>

Industria	<input type="checkbox"/> Manufatti <input type="checkbox"/> Venditori di software <input type="checkbox"/> Venditori di hardware <input type="checkbox"/> Telecomunicazioni <input type="checkbox"/> Internet Service Providers <input type="checkbox"/> Servizi dei motori di ricerca <input type="checkbox"/> Banche e assicurazioni <input type="checkbox"/> Pubblicità <input type="checkbox"/> Dettaglio	Accademico	<input type="checkbox"/> Progetti universitari e accademici <input type="checkbox"/> Subject Gateways <input type="checkbox"/> Iniziative W3C
Editoria	<input type="checkbox"/> Distribuzioni <input type="checkbox"/> Musica/video/multimedia distr <input type="checkbox"/> Industria registrazioni musica <input type="checkbox"/> Riviste scientifiche <input type="checkbox"/> Agenzie stampa <input type="checkbox"/> Quotidiani <input type="checkbox"/> Gestione dei diritti	Educazione e insegnamento	<input type="checkbox"/> Corsi interattivi <input type="checkbox"/> Educazione permanente e continua durante l'arco della vita <input type="checkbox"/> Insegnamento a distanza <input type="checkbox"/> Scuola <input type="checkbox"/> Curricula
Audio-video	<input type="checkbox"/> Industria filmografica <input type="checkbox"/> Produzioni broadcast e archiviazione delle produzioni multimediali	Ricerca	<input type="checkbox"/> Laboratori di ricerca <input type="checkbox"/> Compagnie di ricerca <input type="checkbox"/> Società professionali
Beni culturali	<input type="checkbox"/> Biblioteche <input type="checkbox"/> Servizi bibliotecari <input type="checkbox"/> Musei <input type="checkbox"/> Portali di musei <input type="checkbox"/> Archivi	Informazioni geografiche	<input type="checkbox"/> Fornitori di informazioni geospaziali
		Altri domini	<input type="checkbox"/> Trasporti <input type="checkbox"/> Servizi di spedizione posta elettronica <input type="checkbox"/> Sanità <input type="checkbox"/> Servizi governativi <input type="checkbox"/> Ambiente

Metadati nelle comunità di utenti

11. Chi produce i metadati?

Su questo sfondo si possono notare alcune tendenze, specialmente a cavallo dei confini delle tre fasce nella suddivisione tipologica strutturale della prima tabella.

I metadati prodotti dagli autori o dal sito acquisiranno maggiore importanza per molti scopi. Una motivazione importante è il superamento di alcune delle deficienze degli attuali motori di ricerca, evitando di addossare ai provider i costi della creazione dei record. Per certi versi, i motori assumeranno alcune delle caratteristiche della fascia di mezzo, anche per gli aspetti funzionali. Allo stesso tempo, le comunità che utilizzano i più ricchi formati di "documentazione" saranno interessate a mettere a disposizione di un più vasto pubblico l'informazione sulle loro risorse.

La via migliore per raggiungere questo obiettivo è quella che consente che record funzionali ad attività di "presentazione" siano esportati in altri sistemi. Questo orientamento porta i metadati classificati ora come di seconda fascia a divenire più importanti, a livello funzionale, come via di accesso generalista, senza escludere, in alcuni casi, integrazioni con link a record specifici di dominio.

Comunità di utenze specifiche che producono e parallelamente fruiscono di contenuti intellettuali, partecipano quindi alla creazione di metadati "esponendoli" alle altre comunità, in uno scambio reciproco di informazioni che alloggiato entro i metadati.

In questa fase i metadati sono beni informativi il cui scambio è attuato attraverso "trasporti" tra un luogo e l'altro della Rete. E' il caso per esempio dello scambio tra metadati che avviene

entro la piattaforma Open Archive Initiative, dove i *repositories* di documenti del settore accademico e di ricerca vengono dotati di sistemi per l'estrazione automatica e conseguente "esposizione" di metadati ai fini di una raccolta da parte di motori specifici e indicizzazione entro sistemi.

Gli stessi autori dei documenti sono oggi facilitati nella creazione di metadati da incardinare dentro i documenti da loro prodotti. Al contempo le stesse comunità cui afferiscono i creatori delle risorse si stanno dedicando, soprattutto in certi domini specifici, alla predisposizione di formati standard e, in certi casi anche alla creazione di linguaggi di marcatura specialistici.

Nel mondo degli Open Archive sta destando molto interesse anche la ricerca sull'estrazione automatica di metadati dai documenti a testo pieno.

Una delle conseguenze di questa ampia gamma di comunità che hanno interesse nei metadati è che c'è un sconcertante numero di standard e formati già esistenti o sotto sviluppo³⁹.

Il mondo delle biblioteche, per esempio, ha sviluppato i formati MARC con lo scopo di codificare metadati definiti nelle regole catalografiche e ha anche definito standard descrittivi nelle serie ISBD *International Standard Bibliographic Description* (ISBD).

Altri domini hanno definito standard di metadati ricchi basati sull'implementazione di *Standard Generalised Markup Language* (SGML) o *Extensible Markup Language* (XML) i quali sono meta-linguaggi. Esempi di questo tipo sono i metadati *Encoded Archival Description* (EAD) e i *CIMI Document Type Definition* (DTD).

Tali applicazioni si basano primariamente sull'effettivo recupero degli oggetti rappresentati dai metadati che li descrivono e che sono espressi da parte di una data comunità di utenti.

I bibliotecari dovranno cambiare atteggiamento, per troppo tempo hanno considerato la catalogazione come un loro esclusivo dominio. Riprendo nuovamente Hakala in un passo che mi sembra di fondamentale importanza quando afferma che lasciare che altri esperti creino metadati senza che questi siano necessariamente rivisti dai bibliotecari, "*può essere causa di pena per qualche collega*", ma in termini generali sicuramente sarà di notevole beneficio in quanto solo con l'apporto diretto di autori ed editori sarà possibile giungere ad una descrizione esaustiva degli articoli delle riviste. "*Data l'enorme quantità di metadati che i nuovi sistemi di identificatori richiedono è ovvio che biblioteche, editori ed autori stessi devono unire le forze nel conseguire il risultato. [...] Istituire questa cooperazione trasversale sarà una sfida interessante, sia tecnicamente sia mentalmente.*"⁴⁰

Formati proprietari e formati standard

12. I metadati sono sempre ad accesso libero?

Un aspetto che va considerato, come osserva Lorcan Dempsey, è che molte comunità di soggetti, in numerosi settori di mercato, sono fortemente attaccati ai loro formati. Questi formati proprietari su cui specialisti di vari settori del mercato dell'informazione hanno investito in termini di documentazione, conoscenza e apprendimento, sono stati sviluppati entro i sistemi creati per manipolare i dati e offrire servizi ad utenze specifiche.

Per queste ragioni questa diversità di formati è qualcosa che si perpetua nel tempo e, a dire il vero, nuovi formati di metadati saranno periodicamente sviluppati per soddisfare i bisogni di differenti comunità di soggetti.

³⁹ Michael Day "*Metadata in a nutshell*". UKOLN: the UK Office for Library and Information Networking, University of Bath, UK. [Draft of an article published in *Information Europe* 6(2), Summer 2001, p. 11. *Information Europe* is the quarterly magazine of EBLIDA (the European Bureau of Library, Information and Documentation Associations)].

<<http://www.ukoln.ac.uk/metadata/publications/nutshell/>>

⁴⁰ Juha Hakala, op. cit.

Esiste tensione tra questo continuo progredire di formati specialisti e qualsiasi requisito per un livello di interoperabilità, il quale dovrebbe permettere un'adeguata presentazione delle risorse attraverso i domini di soggetti e i tipi di informazione.

In questa prospettiva va considerato che i formati proprietari adottati negli OPAC venduti da commerciali non è detto si adeguino a formati internazionali indispensabili allo scambio di metadati, innescando meccanismi di copyright sui metadati⁴¹.

Parimenti va detto che, anche quando vi sia adozione di formati catalografici standard, vi possono essere limitazioni di accesso ai metadati dell'OPAC dovute a "licenze sull'interfaccia proprietaria" le quali possono regolare gli accessi in base al numero di utenti contemporanei.

In questi casi non si tratta di utilizzo di metadati proprietari, ma di limitazioni all'utilizzo degli stessi i quali dovrebbero essere un bene collettivo condivisibile, senza restrizioni.

Identificatori, risolutori di nomi e algoritmi per URN entro OPAC aperti, in realtà potrebbero divenire "informazioni di dominio chiuso", laddove applicati a soluzioni proprietarie.

Si potrebbe arrivare al paradosso, per nulla lontano in questi casi, che anche sui metadati catalografici vi possano gravare diritti di proprietà da parte dei fornitori di contenuto⁴².

Interoperabilità tra formati di metadati

13. Qual è la relazione tra metadati e interoperabilità?

L'esistenza di diversità di formati, ma soprattutto la tremenda diversità tra risorse informative che confluiscono entro un singolo sistema globale che è il Web, presentano notevoli implicazioni in termini di interoperabilità. Interoperabilità e metadati sono quindi parole chiave che consentono di aprire e mettere in comunicazione mondi diversi.

L'interoperabilità può essere definita come l'abilità dei sistemi di fornire servizi e di accettare servizi da altri sistemi in termini di comunicazione e di comprensione reciproca attuata attraverso un colloquio reso possibile nonostante le differenze.

La natura volatile ed effimera dei documenti Web e l'incessante aumento di informazione disponibile sta conducendo verso servizi di ricerca alternativi ai tradizionali cataloghi i cui costi sono molto alti. I servizi che incorporano i vantaggi di un approccio automatico per l'indicizzazione si stanno orientando verso l'uso di formati semplici per la descrizione delle risorse. Al contempo l'interoperabilità tra formati di metadati consente un aggancio anche ai dati ricchi e strutturati dei cataloghi delle biblioteche consentendo un recupero dei documenti sia cartacei sia digitali all'interno dei servizi offerti dalle biblioteche ibride.

Una risposta a questo problema è la produzione di *crosswalk mapping* di metadati tra uno o più formati⁴³. Ad un livello di base il *crosswalk* può essere utilizzato ai fini della comparazione tra formati al fine di stabilire il loro potenziale di interoperabilità. In tal senso possono anche essere utilizzati, come base per la produzione di uno specifico programma di conversione di formati o, potenzialmente, per la produzione di sistemi di ricerca, i quali dovrebbero permettere l'interrogazione di formati di metadati eterogenei.

L'interoperabilità in tale contesto richiede perciò l'adozione di un nucleo di formati di metadati, del tipo proposto dalla *Dublin Core Initiative*, al fine di agire come intermediari per l'interoperabilità semantica tra modelli descrittivi di risorse eterogenee.

⁴¹ Querelle Open Metadata Framework versus Copyrighted Content Framework

⁴² Antonella De Robbio "OPAC ricombinanti tra nuovi assetti di mercato e cambiamenti tecnologici"

48. congresso (Roma) AIB2001 3-5 ottobre . Bibliocom 2001. OPAC: punto e a CAPO. Quali cataloghi per il recupero delle informazioni bibliografiche in linea (Seminario AIB-WEB 4)

⁴³ *Mapping between metadata formats*. Pagina sul sito UKOLN a cura di Michael Day che raggruppa la mappatura tra differenti schemi di metadati da o verso il Dublin Core
< <http://www.ukoln.ac.uk/metadata/interoperability/>>

Stu Weibel suggerisce che la promozione di un "insieme comunemente inteso di un nucleo di descrittori migliorerà le prospettive per la ricerca tra discipline diverse attraverso l'unificazione di attributi correlati". Un approccio all'interoperabilità in un ambiente descrittivo di risorse eterogenee dovrebbe essere la mappatura di molti schemi di descrizione entro un insieme comune come il Dublin Core il quale potrebbe dare all'utente "un singolo modello semantico per la ricerca". Una piena interoperabilità sarà possibile soltanto nella relazione di tre requisiti fondamentali. Questi sono: XML/RDF/DC.

XML il linguaggio flessibile

14. Quali prospettive per i metadati?

XML *Extensible Markup Language* è il nuovo metalinguaggio flessibile sviluppato dal *World Wide Web Consortium* e progettato per rendere le informazioni autodescrittive. XML superando le incompatibilità tra sistemi operativi legati all'ambiente hardware e software comporterà una rivoluzione nel modo di comunicare, ampliando alla massima potenza la possibilità di scambiare dati con estrema rapidità.

HTML è un linguaggio rigido, la flessibilità di XML consente invece di utilizzare lo stesso meccanismo dei *tag*, o etichette in modo che ci dicano che cos'è quell'informazione.

Ora le etichette di HTML ci dicono solo "come" una risorsa si presenta, per esempio:

```
<H2>titolo del documento</H2>
```

```
<bold> autore in grassetto</bold>
```

Con XML sarà possibile invece definire entro le etichette "cosa" sono le informazioni che si stanno inviando:

```
<titolo>titolo del documento</titolo>
```

```
<autore>cognome, nome</autore>
```

Con XML si potranno scambiare dati strutturati semplici in modo semanticamente corretto.

XML consiste in regole che chiunque può adottare per creare dal nulla un proprio linguaggio.

Ciò senza dubbio sarà di estremo interesse per esempio per le attività entro domini medici o farmacologici dove la possibilità di uno scambio di dati etichettati in modo adeguato può codificare referti, prodotti, studi.

Applicato alla matematica XML ha già condotto alla creazione di *MathML* un linguaggio specifico che permette di inserire equazioni in pagine Web con poche righe di testo, utile alla comunità dei matematici a lungo penalizzata a causa degli attuali navigatori (browser) Web.

Per i chimici *CML Chemical Markup Language* rende graficamente la struttura molecolare dei composti descritti, per gli astronomi è stato sviluppato *AML Astronomy Markup Language* mentre *Biosequence ML* rivolto alla comunità dei genetisti, in via di sviluppo, servirà ai progetti di mappatura e sequenziamento dei geni.

Per questa sua caratteristica XML comporterà uno sviluppo notevole in molti settori della ricerca ove i metadati saranno al centro di nuovi processi comunicativi, mezzi per comunicare teorie, calcoli e risultati sperimentali. C'è chi afferma che XML sarà la fine del WWW inteso come *World Wide Wait* in quanto migliorerà notevolmente i tempi di risposta, per il fatto che tutte le informazioni strutturali e semantiche aggiunte al metadato con XML, si troveranno direttamente sul client dell'utente, giunte direttamente dal server. Ciò eviterà i pesanti scambi attuali tra client/server di informazioni statiche, riducendo enormemente il traffico di rete.

XML offrirà soluzioni vantaggiose anche a persone con disabilità visive in quanto i fogli di stile *Extensible Stylesheet Language XSL* permetteranno di convertire XML in Braille o di deviarne i contenuti al sintetizzatore vocale.

<XML permetterà non solo di scambiare informazioni tra differenti sistemi operativi, ma anche di superare confini nazionali e culturali./>

<XML e XSL permetteranno agli editori di scrivere una sola volta e pubblicare ovunque, su carta o in forma elettronica./>⁴⁴

Dublin Core, per il mondo bibliotecario, codificato entro XML, allo stato attuale e in ragione della sua struttura semplice, facilmente leggibile dalla macchina, pare essere l'insieme degli elementi chiave che verrà usato nello sviluppo di altri standard.

Inoltre, come suggerisce Hakala⁴⁵, laddove in fase di produzione della risorsa elettronica si utilizzi XML, sarà molto semplice estrarre automaticamente, in una fase successiva, metadati utili ai nostri cataloghi. Ecco quindi come i metadati incorporati in XML, aiuteranno nella gestione dei materiali più importanti ora poco visibili entro i nostri cataloghi di biblioteca. A patto che vi sia cooperazione tra autori, editori e biblioteche.

RDF la terra di mezzo

15. Quando un modello per la rappresentazione dei metadati?

Il nuovo standard per i metadati Web è RDF *Resource Description Framework (RDF)*⁴⁶, un modello per metadati, o meglio una piattaforma per il trattamento dei metadati, il quale fornisce l'interoperabilità tra le applicazioni che scambiano sul Web informazioni comprensibili dalle macchine. Tale modello consente di rappresentare metadati in una sintassi per la codifica e la trasmissione secondo modalità che favoriscano l'interoperabilità fra server e client Web, sviluppati in modo indipendente.

Questo *framework* permette la descrizione di risorse Web per facilitare l'elaborazione automatica delle informazioni e può essere utilizzato in diverse aree di applicazione

- ❑ nella ricerca di risorse al fine di migliorare le capacità dei motori di ricerca,
- ❑ nella catalogazione per descrivere il contenuto e le relazioni del contenuto disponibili in un particolare sito Web,
- ❑ in una pagina o in una biblioteca digitale, usato da agenti software intelligenti per facilitare la condivisione e lo scambio di conoscenza,
- ❑ nella valutazione di contenuto,
- ❑ nel descrivere collezioni di pagine che rappresentano un unico "documento" logico,
- ❑ per descrivere i diritti di proprietà intellettuale di pagine Web,
- ❑ per esprimere le preferenze sulla riservatezza da parte di un utente così come le politiche di riservatezza di un sito Web.
- ❑ associato alla firma digitale rappresenterà la chiave di volta per la costruzione di un "Web affidabile" per il commercio elettronico, per la collaborazione e per altre applicazioni.

RDF non assume nessuna particolare applicazione o dominio semantico in quanto consiste di una descrizione di nodi e coppie attributo/valore attaccate. I nodi possono essere qualsiasi risorsa Web, che è un URI, il quale include l'URL. Gli attributi sono proprietà dei nodi e i loro valori sono stringhe di testo o altri nodi (risorse Web o istanze di metadati).

Il modello dei dati RDF offre una struttura concettuale, astratta, per la definizione e l'uso di metadati. Per creare e scambiare questi metadati è necessaria anche una sintassi concreta.

A grandi linee la sintassi di RDF usa XML all'interno di un modello che poggia su tre concetti base: *il soggetto, il predicato, l'oggetto*.

1. Il *soggetto* è la risorsa descritta.

⁴⁴ John Bosak e Tim Bray, op. cit

⁴⁵ Juha Kakala, op. cit.

⁴⁶ *RDF (Resource Description Format)* <<http://www.w3.org/RDF/>>

L'*RDF Resource Description Framework* è descritta nel documento W3C "Resource Description Framework (RDF) Model and Syntax Specification" all'indirizzo <<http://www.w3.org/TR/PR-rdf-syntax/>>. Dal testo originale sul sito W3C si sono riprese delle parti, tradotte e in parte rielaborate

2. Il *predicato* è la caratteristica o l'aspetto specifico utilizzato per la descrizione del soggetto, che determina la proprietà della risorsa descritta.
3. L'*oggetto* è quell'affermazione che attribuisce ad una caratteristica della risorsa descritta un determinato valore.

L'elemento RDF è un semplice involucro che delimita i confini in un documento XML entro cui il contenuto è esplicitamente da intendere mappabile in un'istanza di modello di dati RDF. RDF è complementare a XML in quanto una delle finalità di RDF è di permettere la definizione di semantiche per dati in formato XML, secondo modalità standardizzate e che mirano all'interoperabilità.

Ci si aspetta che RDF nel prossimo futuro influenzerà le iniziative di standardizzazione che sono attualmente ancora in fase di sviluppo o quelle che sono appena iniziate

Metadati chiave di accesso alla biblioteca ibrida

16. Metadati perché?

Insiemi di metadati formano cataloghi che possono risiedere in luoghi indipendenti rispetto agli oggetti che essi descrivono. Così accade nelle biblioteche tradizionali dove, da sempre, gli oggetti fisici abitano scaffali, mentre i cataloghi, con i record che li rappresentano, risiedono altrove, anche fuori della biblioteca fisica.

In questo caso si tratta di metadati esterni ai documenti che, se opportunamente organizzati, vanno a comporre insiemi informativi - più o meno specialistici - che si configurano come veri e propri cataloghi e, in senso più ampio, come servizi di accesso all'informazione.

La biblioteca ibrida è composta quindi da:

1. collezioni di metadati che si riferiscono a documenti su supporti tradizionali: OPAC, banche dati bibliografiche, archivi, ...
2. collezioni di metadati che si riferiscono ad oggetti digitali: SBIG *Subject-Based Information Gateway*, archivi digitali, cataloghi di risorse elettroniche ...

Le collezioni di metadati di materiale tradizionale descrivono solitamente oggetti su supporto fisico: i contenuti degli scaffali di biblioteche, gli oggetti delle mediateche, i documenti d'archivio, le fotografie, ... Sono solitamente oggetti del mondo reale delle cose, come libri, riviste su carta, CD-ROM; cassette, video, filmati, fotografie, lettere,.... tutti oggetti che sovente sopravvivono agli uomini stessi e che, incarnati in supporti più o meno durevoli nel tempo, sono espressione delle loro idee. Tali oggetti abitano spazi fisici definiti.

Le collezioni di metadati della biblioteca digitale si riferiscono invece a oggetti del mondo virtuale. Questi oggetti digitali possono essere la trasposizione di oggetti che abitano il mondo reale delle cose, per esempio la digitalizzazione di un testo o la conversione di una rivista da cartacea a digitale, o la digitalizzazione di foto e documenti d'archivio o di video e audio.

In questo caso si ha corrispondenza tra i due mondi.

Esistono però oggetti che vivono solo nella dimensione digitale e che non hanno corrispettivo nel mondo reale delle cose: siti Web, repertori, risorse e documenti in rete, gli archivi di documenti digitali a testo pieno, i server *streaming* per la multimedialità digitale, i *repositories* della ricerca, le riviste elettroniche nate digitali, gli *e-book*, gli archivi di immagini digitali, ... Sono tutti oggetti che, in virtù del loro essere effimero, potrebbero anche non sopravvivere al mondo delle cose reali, se non opportunamente sottoposti a un retaggio che li trasponga da uno spazio ad un altro spazio, da un tempo ad un altro.

La pervasività dei metadati ha già cominciato a stravolgere modelli e architetture, sia entro le mura fisiche sia nel ciberspazio.

L'integrazione tra i due mondi avviene attraverso i portali della biblioteca ibrida, che poggiano su architetture di metadati mobili.

I metadati allora divengono la cerniera tra il vecchio mondo e il nuovo, tra la solida tradizione e la tecnologia che avanza. Utilizzati per nuove forme di catalogazione, i metadati sono la chiave che consentirà di aprire i cancelli di ciò che è chiamato oggi controllo bibliografico e che potrebbe sconfinare, molto presto, verso nuovi contesti, assai più ampi, nella gestione dell'informazione di rete, dentro e fuori le mura della biblioteca ibrida.