

## Metadattazione e interoperabilità

SILVIA PANZAVOLTA<sup>1</sup>

*Il testo affronta la complessa questione dell'architettura dell'informazione in Internet, con particolare riferimento alla trovabilità dei dati e all'interoperabilità tra schemi di descrizione (metadati) di diversi sistemi di informazione, sia a livello nazionale sia a livello internazionale. Nell'intento di offrire una panoramica delle varie problematiche, soluzioni e standard che la comunità degli architetti dell'informazione online sta trattando, l'analisi si concentra sul settore disciplinare delle scienze dell'educazione.*

Parole chiave: Documentazione educativa - Sistema di informazione - Metadati - Standard - Interoperabilità - Trovabilità

I metadati (o *metadati*), ossia dati che descrivono altri dati, sono recentemente oggetto di grande attenzione. Dal 1991, anno in cui Tim Berners Lee<sup>2</sup> sviluppò il *World Wide Web*, costruito prevalentemente in HTML, tutti, e non solo una minoranza di esperti e tecnici, sono in grado di consultare e pubblicare informazioni in rete. Così, da quel momento in poi, Internet ha visto una crescita esponenziale, come si può osservare dalla figura 1 (trad. it.) tratta dal sito Search Engine Watch<sup>3</sup>, in cui vengono riportati i bilioni di pagine indicizzate dai motori di ricerca più frequentati (Google, Yahoo ecc.).

Figura 1 - Bilioni di pagine indicizzate dai principali motori di ricerca  
(fonte: *Search Engine Watch*, trad. it.)

<sup>1</sup> Silvia Panzavolta collabora con Indire [Istituto Nazionale di Documentazione per l'Innovazione e la Ricerca Educativa] come documentalista *senior* alla gestione di progetti di ricerca nazionali e internazionali. Per contattarla, scrivere a <s.panzavolta@indire.it>.

<sup>2</sup> <<http://www.w3.org/People/Berners-Lee/>>.

<sup>3</sup> <<http://searchenginewatch.com/>>.

Anche il numero dei domini ha subito un'impennata, in particolare nell'ultimo quinquennio, come mostrato nel grafico in figura 2, riportato sul sito dell'Internet Systems Consortium<sup>4</sup>.

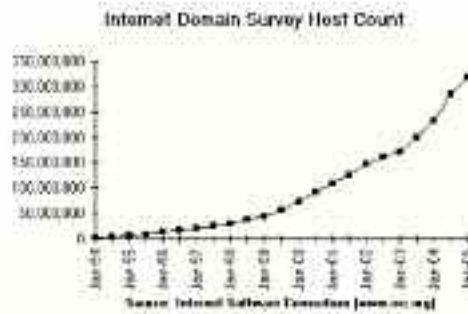


Figura 2 - Grafico relativo all'andamento dei domini Internet dal 1994 ad oggi

Risulta evidente come una tale mole di informazioni debba essere gestita e organizzata in modo "logico", per poi poterla ricercare. Alcune nuove figure professionali, come l'*architetto delle informazioni*, il *bibliotecario dell'informazione digitale* o il *newsmaster*, testimoniano la grande richiesta di competenze nel settore dell'organizzazione e della ricerca dei bilioni di dati su Web.

È opportuno evidenziare che l'informazione in Internet è diversamente stratificata nei siti, tanto che si parla di *deep web* (web profondo), detto anche *invisible web* (Sherman, 2001), e di *surface web*. Il primo sarebbe il web costituito da archivi elettronici, banche dati, repertori ecc. che i normali motori di ricerca non riescono a "vedere", ma in cui le informazioni sono strutturate e il recupero dei dati produce poco rumore, ossia risposte non pertinenti. Il *surface web*, invece, è costituito da pagine HTML interconnesse tra loro da *link* ipertestuali. Mentre il *surface web* è composto da una rete di rimandi, benché non strutturati, il *deep web*, che secondo una stima è circa 60 volte più consistente (figura 3) del web superficiale (circa 500 miliardi di documenti, Bright Placet, 2000), raramente presenta interconnessioni tra le varie banche dati e, perfino all'interno di uno stesso sito, raramente l'amministratore prevede una comunicazione tra i vari archivi.

<sup>4</sup> <<http://www.isc.org/index.pl>>.

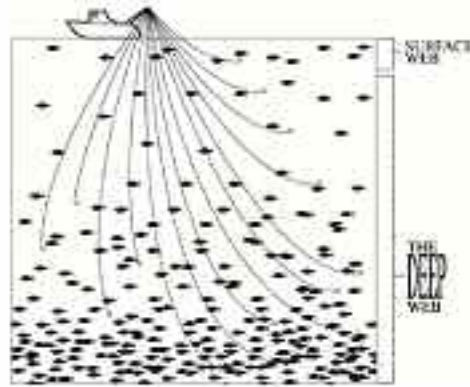


Figura 3 - Rapporto tra *deep web* e *surface web* (fonte: *MotoriDiRicerca.IT*)

Dunque, se da una parte tale organizzazione dell'informazione all'interno di una banca dati consente una migliore precisione dell'*information retrieval*, cioè del recupero dell'informazione, dall'altra essa potrebbe non essere esaustiva in quanto l'informazione è frammentata in banche dati diverse e non comunicanti. Si tratta del problema dell'interoperabilità. È bene distinguere tra interoperabilità tecnica e interoperabilità concettuale. La prima riguarda l'interoperabilità tra ambienti di apprendimento *online*, tra *software* di gestione dei contenuti ecc. e riguarda gli aspetti informatici del problema. L'interoperabilità concettuale riguarda, invece, gli aspetti di significato (semantica) e quelli rappresentazionali (semiotica).

Abbiamo, dunque, toccato i due principali aspetti dell'architettura dell'informazione su web: la descrizione dei dati (metadati) e l'interoperabilità.

È vero che alcune banche dati sono comunque ricercabili a partire da un'unica maschera di ricerca (si pensi alle iniziative dell'ICCU - Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche per la creazione di un catalogo unico delle biblioteche italiane, cui aderiscono 2.382 centri), ma si tratta di soluzioni *a posteriori* e non sempre (come invece accade per l'ICCU) si condivide lo stesso modello descrittivo e/o si documenta lo stesso oggetto (ad es. banche dati bibliografiche, banche dati di immagini, banche dati di *software* ecc.).

Poiché l'informazione all'interno delle banche dati e degli archivi, diversamente dalle pagine HTML sul web, è organizzata, è possibile scegliere su quale elemento descrittivo dell'oggetto fare ricerca (ad es. Titolo, Autore, Abstract o Descrizione, Soggetto). Bisogna precisare che le pagine HTML potrebbero contenere i metadati nella parte "head" della pagina, ma di fatto quasi nessuno se ne serve. Inoltre, i metadati che generalmente vengono inseriti non sono in alcun modo standardizzati e contengono descrizioni limitate. Infine, il linguaggio HTML pare inadeguato a descrivere in modo semanticamente significativo il contenuto di una pagina web.

Per cercare di ovviare a questa idiosincrasia e tentare una qualche struttura dell'informazione *online*, vari enti, iniziative e progetti interessati al tema stanno proponendo di utilizzare un linguaggio tale da poter essere letto dai motori di ricerca e da fornire informazioni sugli elementi descrittivi dell'oggetto: l'XML, eXtensible Mark-up Language (cfr. *XML in 10 punti*<sup>5</sup>). In tal modo, sarebbe possibile sia strutturare un'intera banca dati perfettamente ricercabile dai motori di ricerca (i cui record potrebbero anche essere esportati da una banca dati all'altra o "letti" da altri archivi), sia avere un livello descrittivo semantico nelle singole pagine web.

I metadati possono essere applicati a qualsiasi risorsa (elettronica e non) e di fatto sono assimilabili alle schede catalografiche di una biblioteca. La prima iniziativa che si è occupata della descrizione dei dati su web è la DDCMI [Dublin Core Metadata Initiative]<sup>6</sup> che, nel 1999, ha rilasciato la versione 1.1 del Set di elementi di metadattazione Dublin Core, che prevede una descrizione della risorsa in base a 15 elementi, declinati secondo 10 attributi:

- 1) Titolo (*Title*): il titolo o il nome della risorsa, assegnato dall'autore;
- 2) Autore (*Creator*): l'autore (persona o ente) primariamente responsabile per la realizzazione della risorsa;
- 3) Autore di contributo secondario (*Contributor*): la responsabilità secondaria dell'opera, sia esso il Curatore, l'Illustratore, il Traduttore, ecc.;
- 4) Editore (*Publisher*): il soggetto (persona, ente, servizio) responsabile della pubblicazione della risorsa;
- 5) Data (*Date*): la data di creazione o di pubblicazione della risorsa; il formato è quello regolato dalla norma ISO 8601, per cui la data va espressa con AAAA-MM-GG (per es. 2005-08-09);
- 6) Soggetto (*Subject*): l'argomento o gli argomenti trattati nella risorsa; questo campo consente di effettuare ricerche semantiche a partire da categorie concettuali, concetti (parole chiave) o codici di classificazione (per es. Classificazione Decimale Dewey), a seconda dello schema di riferimento adottato;
- 7) Descrizione (*Description*): una descrizione testuale della risorsa espressa in linguaggio naturale (può essere un *abstract*, l'indice dei contenuti, una recensione ecc.);
- 8) Tipo (*Type*): la natura o il genere della risorsa - è raccomandato il vocabolario controllato messo a punto da Dublin Core, che prevede: raccolta/collezione; set di dati strutturati (per es. banca dati); evento (conferenza, seminario, esibizione); immagine; risorsa interattiva (include: *applet*, *learning object* multimediali; *chat*; realtà virtuali); immagine in movimento (per es. film, animazione, video ecc.); oggetto fisico, ossia un'entità non animata e tridimensionale (per es. una piramide, una statua); ser-

<sup>5</sup> <<http://www.indire.it/content/index.php?action=read&id=313>>.

<sup>6</sup> <<http://dublincore.org/documents/demi-terms/>>.

vizio, ossia qualsiasi tipo di servizio pensato per un utente finale (per es. un servizio di fotocopie, un servizio bancario o un servizio di prestito interbibliotecario); *software* (programma informatico che può essere installato da un computer ad un altro); suono (un *compact disc*, un file audio, la registrazione di un'intervista); immagine ferma (una mappa, un disegno o un dipinto); testo (tutte le risorse che contengono prevalentemente testo da leggere, per es. articoli, poesie, dossier, libri, *e-mail* ecc.);

9) Formato (*Format*): la manifestazione della risorsa, sia essa fisica o digitale. Viene usato per esprimere il formato e le dimensioni (durata e grandezza). Anche per questo campo si suggerisce di basarsi su un vocabolario controllato, il MIME Media Type<sup>7</sup> messo a punto da IANA [Internet Assigned Numbers Authority];

10) Identificatore (*Identifier*): un riferimento non ambiguo alla risorsa nell'ambito di un dato contesto (per es. l'URI, Uniform Resource Identifier della risorsa oppure un numero standard, come l'ISBN, International Standard Book Number o il DOI, Digital Object Identifier);

11) Lingua (*Language*): la lingua del contenuto intellettuale della risorsa, secondo la norma ISO 639-1 e 639-2 (2 o 3 cifre);

12) Relazione (*Relation*): riferimento ad una risorsa con un forte legame di implicazione con la risorsa descritta; meglio se tale riferimento viene espresso attraverso una stringa non ambigua (URI, ISBN, DOI ecc.);

13) Copertura (*Coverage*): gli elementi spaziali o temporali relativi al contenuto della risorsa (per es. Firenze<sup>8</sup>, Medioevo ecc.). Anche per questo campo si raccomanda l'uso di un vocabolario controllato, nella fattispecie il Getty Thesaurus of Geographic Names Online<sup>9</sup>;

14) Fonte (*Source*): informazioni sulla risorsa dalla quale la risorsa descritta è stata derivata; meglio se tale riferimento viene espresso attraverso una stringa non ambigua (URI, ISBN, DOI ecc.);

15) Diritti (*Rights Management*): dichiarazione circa la gestione dei diritti o riferimenti relativi ad un servizio/ente che fornisce tale informazione.

Quelli finora presentati sono gli elementi di base, ai quali si affiancano dati più specifici. Si può osservare infatti che la scheda Metadata Dublin Core richiama una scheda catalografica e dice molto poco sull'uso che di tale risorsa si può fare, in particolare in ambito educativo. Alcuni elementi sono stati successivamente aggiunti, proprio per sopperire a questa mancanza. La DCMI prevede, infatti, l'aggiunta di

<sup>7</sup> <<http://www.iana.org/assignments/media-types/>>.

<sup>8</sup> <[http://www.getty.edu/vow/TGNHierarchy?find=florence&place=city&nation=italy&prev\\_page=1&english=Y&subjectid=7000457](http://www.getty.edu/vow/TGNHierarchy?find=florence&place=city&nation=italy&prev_page=1&english=Y&subjectid=7000457)>.

<sup>9</sup> <[http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/index.html](http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html)>.

elementi come il livello educativo del destinatario (*Education Level*) e il modello di apprendimento a cui la risorsa si ispira (*Instructional Method*), ma si tratta di un adattamento, in quanto lo schema Metadata Dublin Core nasce per descrivere *online* qualsiasi risorsa (e non per descrivere qualsiasi risorsa *online*!).

Il modello Dublin Core, con alcune modifiche e integrazioni, è stato adottato come standard all'interno del progetto europeo ETB (European Treasury Browser), finalizzato alla costruzione di un *repository* (repertorio) europeo di risorse educative. Il modello metadati ETB, ad esempio, fornisce vocabolari controllati per la tipologia di risorsa educativa (per es. risorsa su attività curricolare; risorsa su attività extra-curricolare; progetto educativo; periodico/giornalino scolastico; *software* educativo; risorsa per la formazione dell'insegnante ecc.), l'età del destinatario ed il contesto di apprendimento (istruzione assistita dal computer, apprendimento cooperativo, apprendimento per scoperta, metodo Montessori, ecc.).

Consulta lo schema Metadata di ETB

Lo schema che intende descrivere risorse educative, ed in particolare i *learning object*, è il LOM (Learning Object Metadata - versione 1.3, 2002)<sup>10</sup> a cura di IEEE [Institute of Electrical and Electronics Engineers]<sup>11</sup> (figura 4).



Figura 4 - Rappresentazione concettuale del modello *metadata* LOM  
(fonte: *Sito di IMS Global Learning Consortium*, 2004)

<sup>10</sup> <[http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)>.

<sup>11</sup> <<http://www.ieee.org/portal/site>>.

Il modello *metadata* è ovviamente molto più complesso di quello Dublin Core perché non solo vuole offrire una griglia descrittiva della risorsa, ma anche indicazioni su come dovrebbe essere fruita, sulle caratteristiche dei destinatari e sul paradigma educativo di riferimento. Il LOM è composto da 9 aree descrittive, per un totale di circa 70 elementi descrittivi:

1) categoria Generale (*General*) (11 campi): racchiude le informazioni generali sul *learning object*, come ad esempio un identificatore non ambiguo (URI, ISBN, DOI ecc.), il titolo, la lingua del LO, la descrizione, la copertura temporale e spaziale e la struttura dell'oggetto (atomica; collezione; reticolare; gerarchica; lineare);

2) categoria Ciclo di vita (*Life Cycle*) (6 campi): racchiude le informazioni sulla vita del LO, come ad esempio la versione, i contributi primari e secondari, la data di creazione, pubblicazione ecc.;

3) categoria Meta-metadata (*Meta-metadata*) (9 campi): fornisce informazioni sullo schema Metadata adottato, l'autore/gli autori dello standard, la lingua dello schema (che può essere diversa dalla lingua dell'oggetto), il formato ecc.;

4) categoria Tecnica (*Technical*) (12 campi): contiene informazioni sui requisiti e le caratteristiche tecniche del LO (formato, grandezza, requisiti tecnici, durata ecc.);

5) categoria Educativa (*Educational*) (11 campi): questa categoria è quella di maggior interesse per la comunità di educatori, insegnanti, *instructional designer* ecc., che dalla consultazione di questa sezione possono derivare molte informazioni utili in merito all'uso educativo dell'oggetto. È utile soffermarsi sugli elementi descrittivi contemplati, ossia:

- Il tipo di interattività (campo 5.1): lo schema distingue tra LO che supportano un tipo di apprendimento attivo, espositivo o misto;
- Il tipo di risorsa educativa (campo 5.2): va indicata la tipologia di oggetto a partire da un vocabolario controllato (esercizio, simulazione, dato, grafico, lezione, diapositiva, questionario ecc.);
- Il livello di interattività (campo 5.3): il LOM presenta 5 livelli, da molto basso a molto alto, con momenti intermedi;
- La densità semantica (campo 5.4): misurata attraverso la durata (file audio e video) o la grandezza, è indipendente dalla difficoltà di fruizione dell'oggetto e presenta 5 livelli (da molto bassa a molto alta);
- Il destinatario primario della risorsa (campo 5.5): il LOM prevede solo 4 tipologie di utenti (insegnante, autore, studente, amministratore);
- Il contesto educativo per il quale il LO è pensato (campo 5.6), ad esempio l'istruzione superiore, l'istruzione obbligatoria, l'addestramento, la formazione professionale ecc.;
- L'età dell'utente tipico (campo 5.7): nel modello LOM, l'età viene indicata

specificando l'età massima e minima dell'utente, per es. 7-9 indica che l'oggetto è adatto a ragazzi dai 7 ai 9 anni;

- La difficoltà dell'oggetto per l'utente tipico (campo 5.8): anche per la difficoltà, si prevedono 5 livelli di difficoltà, da molto facile a molto difficile;
- Il tempo di apprendimento tipico (campo 5.9): il tempo ideale che un utente tipico impiega a fruire l'oggetto;
- La descrizione (campo 5.10): in questo campo l'autore dell'oggetto o chi se ne è servito in modo sistematico (ad es. insegnante), fornisce alcuni suggerimenti o linee guida su come avvalersene;

6) categoria Diritti (*Rights*) (3 campi): fornisce informazioni sui diritti intellettuali, i diritti di riproduzione e le condizioni di uso (e riuso!) dell'oggetto (3 campi);

7) categoria Relazioni (*Relation*) (7 campi): fornisce indicazioni circa il legame tra l'oggetto e altri oggetti o risorse. Poiché il LO è autoconsistente per definizione, ogni rimando dall'oggetto ad un'altra risorsa è bandito. Allora, una strategia è quella di indicare nei metadati (e nel LOM in questa categoria) le relazioni con altri oggetti o risorse;

8) categoria Annotazioni (*Annotation*) (3 campi): contiene commenti sull'uso educativo dell'oggetto e informazioni su autore e data del commento.

Osserviamo che il modello LOM è un modello di metadati interattivo. Mentre l'autore o il documentalista (ma il dibattito su chi debba compilare i metadati è aperto!) inserisce le informazioni nelle altre categorie, questa è riservata a utilizzatori, valutatori, sperimentatori ecc. La scheda *metadata*, allora, lungi dall'essere uno strumento statico, diventa uno strumento dinamico e collaborativo in grado di mettere in collegamento membri (ad es. insegnanti) di una stessa comunità.

Ed infine:

9) categoria Classificazione (*Classification*) (8 campi): contiene informazioni circa il soggetto o la materia curricolare affrontata nel LO. Si possono anche introdurre parole chiave libere, ma bisogna specificare il contesto semantico di riferimento, ad esempio la Classificazione Decimale Dewey (DDC), la Library of Congress Classification (LOC) o il Thesaurus Europeo dell'Educazione (EET o TEE)<sup>12</sup>.

Il modello SCORM [Sharable Content Object Reference Model]<sup>13</sup>, messo a punto nel 2000 dal Dipartimento della Difesa e dal Dipartimento del Lavoro degli Stati Uniti, nell'ambito del progetto ADL [Advanced Distributed Learning]<sup>14</sup>, mira

<sup>12</sup> <[http://www.eurydice.org/TeeForm/FrameSet\\_EN.htm](http://www.eurydice.org/TeeForm/FrameSet_EN.htm)>.

<sup>13</sup> <<http://www.adlnet.org/scorm/index.cfm>>.

<sup>14</sup> <<http://www.adlnet.org/>>.



prevalentemente ad un'interoperabilità tecnica dei LO tra diverse piattaforme di apprendimento (LMS, Learning Management Systems e LMCS, Learning Content Management Systems). Nelle specifiche SCORM, lo schema Metadata adottato per la descrizione dei LO è il LOM.

Il progetto europeo CELEBRATE [Context eLearning with Broadband Technologies]<sup>15</sup>, inizio 2002-fine 2004, coordinato da European Schoolnet<sup>16</sup> e finanziato all'interno del Programma per le Tecnologie della Società dell'Informazione (IST) della Commissione europea, ha coinvolto 23 partecipanti tra Ministeri dell'Istruzione, editori multimediali, sviluppatori, ricercatori ed università di 11 Paesi europei. Il modello *metadata* proposto dal *Metadata Group* (al quale l'autrice ha preso parte) si ispira al LOM, anche se introduce alcuni elementi distintivi, come ad esempio l'adozione di un linguaggio controllato per l'indicizzazione semantica dell'oggetto, ossia il Thesaurus ETB<sup>17</sup> (tradotto ormai in 15 lingue).

Questo strumento (insieme ad altri vocabolari messi a punto all'interno del progetto ETB) è stato indicato dall'IMS Global Learning Consortium<sup>18</sup> come *best practice* di interoperabilità concettuale tra architetture dell'informazione diverse.

All'interno del sistema ETB (figura 5), infatti, sono state messe in comunicazione banche dati diverse, ricercabili a partire dallo stesso strumento (il thesaurus, appunto) grazie ad una precedente operazione di *mapping* concettuale. Le banche dati "mappate" avevano una struttura molto diversa, usavano linguaggi descrittivi diversi ed erano in diverse lingue europee (tedesco, italiano, inglese, danese, spagnolo).



Figura 5 - *Screen shot* del portale ETB (2002), non più *online*

<sup>15</sup> <[http://celebrate.eun.org/eun.org2/eun/en/index\\_celebrate.cfm](http://celebrate.eun.org/eun.org2/eun/en/index_celebrate.cfm)>.

<sup>16</sup> <<http://www.eun.org/portal/index.htm>>.

<sup>17</sup> <<http://www.en.eun.org/eun.org2/eun/en/etb/content.cfm?lang=en&cov=7208>>.

<sup>18</sup> <<http://www.imsproject.org/>>.

Poiché pare superfluo illustrare lo schema metadata CELEBRATE, avendo già ampiamente parlato del LOM, si fornisce la versione italiana <[http://www.indire.it/db/docsrv/Doc/modello\\_metadata\\_CELEBRATE\\_IT.doc](http://www.indire.it/db/docsrv/Doc/modello_metadata_CELEBRATE_IT.doc)> dello stesso, tradotto a cura di Simona Baggiani, Valeria Biggi, Isabel de Maurissens, Franca Pampaloni e Silvia Panzavolta.

Come si potrà osservare, lo schema LOM (così come il modello CELEBRATE) per i *learning object* è molto complesso (circa 70 campi!) e quindi è uno standard riconosciuto ma ancora poco applicato, anche per l'incertezza cui si accennava prima su chi debba compilare la scheda (nel caso di un libro, ad esempio, è il bibliotecario che si occupa della scheda catalografica). Per un'indagine sull'uso del LOM in ambito internazionale, si consulti il *report*<sup>19</sup> a cura del Gruppo di lavoro SC36/WG4<sup>20</sup>, una sottocommissione del *Joint Technical Committee* dell'International Standard Organization (ISO) e dell'International Electrotechnical Commission (IEC).

Ma i problemi non finiscono qui. Alcuni campi, come mostrato, lasciano all'amministratore la libertà di scegliere alcuni vocabolari controllati (ad es. il sistema di classificazione). Dunque, anche se due istituti adottassero lo stesso standard LOM, potrebbero comunque sussistere problemi di interoperabilità (si tratta dell'interoperabilità concettuale, mentre quella tecnica meriterebbe una trattazione a parte).

Per garantire l'interoperabilità, la macchina deve "capire" che tra un sistema descrittivo e un altro c'è una certa relazione (per es. equivalenza, quasi equivalenza o nessuna equivalenza). La soluzione a questa difficoltà è quella di mettere in relazione (intellettualmente) le diverse parti della struttura della scheda (elementi, attributi, vocabolari), sia che corrispondano (in tutto o in parte) sia che non corrispondano.

Facciamo un esempio. Due scuole vogliono condividere *online* il loro catalogo delle risorse didattiche e dei LO. Supponiamo che abbiano adottato entrambe il LOM ma che abbiano deciso di avvalersi di due sistemi di classificazione concettuale diversi. La prima ha scelto la Classificazione Decimale Dewey (CDD) e l'altra il Thesaurus Europeo dell'Educazione (TEE). Se si deve prevedere una ricerca per soggetto, come farà la macchina a cercare contemporaneamente nelle due banche dati? La soluzione che consente la migliore interoperabilità tra i due sistemi è il *mapping*. Bisogna, cioè, analizzare i due sistemi "classificatori" e creare relazioni tra i concetti presenti, indipendentemente dalla loro forma. Già da questo primo esempio possiamo comprendere che il processo non è semplice. La cosa si complica se ci fossero 3 scuole, e non 2, a voler condividere il materiale *online*, e la difficoltà cresce con l'au-

<sup>19</sup> <<http://jtc1sc36.org/doc/36N0871.pdf>>.

<sup>20</sup> <<http://jtc1sc36.org/>>.

mentare del numero degli attori interessati. Figuriamoci, poi, se le scuole fossero anche di nazionalità diversa! In questo caso, avremmo anche il problema del multilinguismo. Ovviamente, l'adozione di uno stesso standard descrittivo (ad esempio il LOM) aiuta, anche se non esaurisce i problemi di interoperabilità concettuale. Il problema è che spesso ogni banca dati è un sistema a sé, magari frutto di una progettazione creativa degli amministratori, ma che finisce per restare un prodotto isolato. Seguire una certa standardizzazione è ormai necessario, anche se pochi sembrano rendersene veramente conto e anzi la percepiscono come una perdita di libertà.

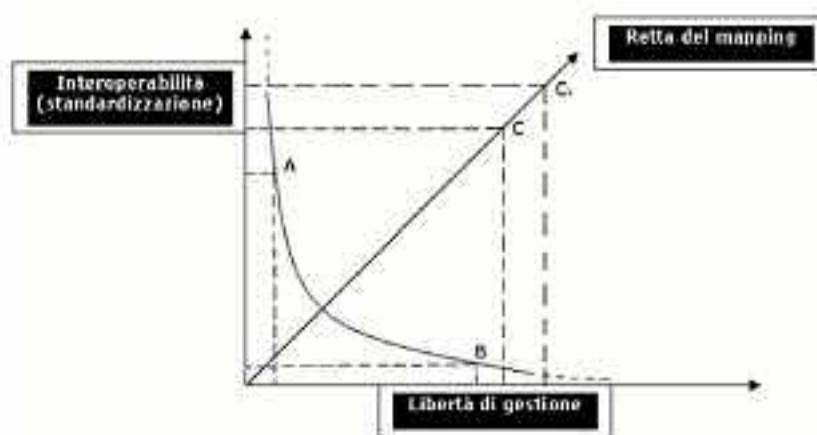


Figura 6 - Rapporto tra standardizzazione e libertà di gestione  
(Autore: Silvia Panzavolta)

In figura 6 viene rappresentato il rapporto tra standardizzazione (interoperabilità) e libertà di creazione/gestione. Come si può osservare, al crescere della standardizzazione (punto A), la libertà decresce e viceversa (punto B). La retta centrale, che rappresenta il *mapping*, garantisce un buon rapporto tra i due aspetti, interoperabilità e libertà di gestione/descrizione (punto C). Il grafico mostra anche che maggiore è la precisione del *mapping*, maggiore sarà sia l'interoperabilità tra sistemi sia il rispetto del modello locale adottato (punto C1).

Il progetto ELEONET [European LEarning Object NETwork]<sup>21</sup>, cui Indire [Istituto Nazionale di Documentazione per l'Innovazione e la Ricerca Educativa] partecipa in qualità di partner, intende sviluppare un modello *metadata* per i *learning object* che non limiti la libertà degli sviluppatori, amministratori e architetti

<sup>21</sup> <[http://www.cineca.it/gai/pagina-eleonet\\_progetto.htm](http://www.cineca.it/gai/pagina-eleonet_progetto.htm)>.

dell'informazione ma che garantisca un alto livello di interoperabilità tra repertori di LO che usano modelli *metadata* diversi. In figura 7 viene rappresentato il modello concettuale ELEONET che non prevede, dunque, un *mapping* tra lo schema A e lo schema B, tra B e C e poi tra A e C (figura 7a), ma un *mapping* di A, B, C, con uno strumento di snodo (figura 7b), l'*index Data Dictionary*<sup>22</sup>, che serve come riferimento centrale e che fa risparmiare il numero di mappature necessarie per rendere interoperabili modelli diversi.

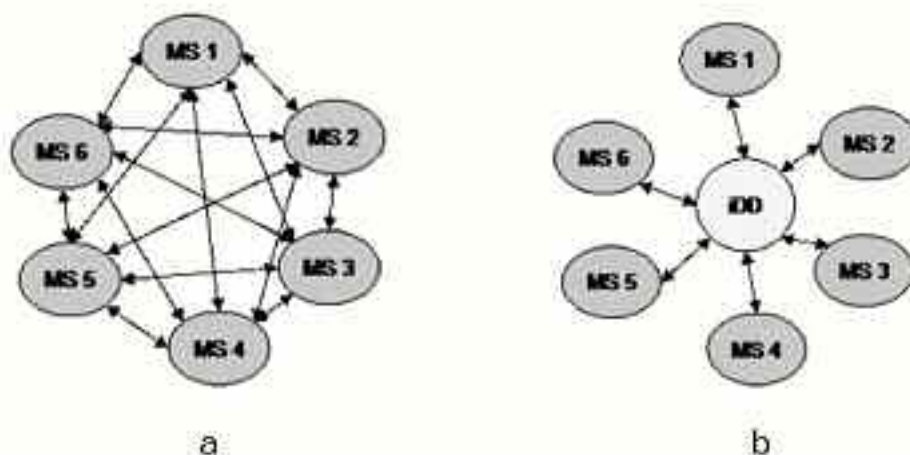


Figura 7 - *Mapping* tra schemi *metadata* diversi (Metadata Schemes, MS), senza l'utilizzo di uno strumento di snodo (disegno a) e con l'utilizzo di uno strumento di snodo (disegno b). Risulta evidente che la soluzione b è molto più veloce ed economica della soluzione a. Infatti, in a il numero dei *mapping* è pari a  $n(n-1)/2$ , ossia 15, mentre in b si avrebbero solo 6 mappature.

In tal modo, dunque, lo schema *metadata* locale può essere conservato in quanto l'interoperabilità è garantita dal *mapping*.

Alla 68<sup>a</sup> Conferenza generale dell'IFLA [International Federation of Library Associations]<sup>23</sup>, tenutasi a Glasgow (Scozia) nel 2002, intitolata *Libraries for life: democracy, diversity, delivery*, il Gruppo di lavoro su classificazione e indicizzazione ha presentato un documento dal titolo *Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis*<sup>24</sup>. Il docu-

<sup>22</sup> <[http://www.doi.org/handbook\\_2000/appendix\\_4.html#A4.6](http://www.doi.org/handbook_2000/appendix_4.html#A4.6)>.

<sup>23</sup> <<http://www.ifla.org/IV/ifla68/>>.

<sup>24</sup> <<http://www.ifla.org/IV/ifla68/papers/008-122e.pdf>>.

mento contiene una rassegna dei lavori di mappatura tra vocabolari diversi, talvolta anche in lingue diverse, effettuati da varie università, biblioteche e centri di documentazione e ricerca. Fino ad ora, sono stati realizzati *mapping* tra vocabolari controllati nella stessa lingua (per es. tra il Library of Congress Subject Headings, LCSH e il Medical Subject Headings) e in lingue diverse (per es. il *mapping* tra i vocabolari usati nel catalogo di tre biblioteche nazionali, quella inglese, quella francese e quella tedesca), tra vocabolari controllati e sistemi di classificazione (per es. tra il Library of Congress Subject Headings, LCSH e la Classificazione Decimale Dewey, CDD) e tra sistemi di classificazione diversi (per es. tra il sistema di classificazione della biblioteca nazionale svedese e la CDD). Nell'ambito del progetto europeo ETB, come anticipato, sono stati effettuati numerosi *mapping* tra il thesaurus ETB e sistemi di classificazione locale o tra il thesaurus ETB e altri thesauri (figura 8).

```
#####
# Mapping of thesaurus
#####
dc:subject:13th century => local:subject:storia moderna
dc:subject:16th century => local:subject:storia moderna
dc:subject:17th century => local:subject:storia moderna
dc:subject:18th century => local:subject:storia moderna
dc:subject:19th century => local:subject:ottocento
dc:subject:20th century => local:subject:avvenimenti storici del xx secolo
dc:subject:21st century => local:subject:novocento
dc:subject:abstraction => local:subject:astrazione
dc:subject:accessories => local:subject:abbigliamento e accessori
dc:subject:acoustics => local:subject:acustica
dc:subject:actor => local:subject:attore
dc:subject:adolescence => local:subject:adolescenza
dc:subject:adolescent => local:subject:ragazzi
dc:subject:adult => local:subject:adulti
dc:subject:adult-child relationship => local:subject:rapporto bambino adulto
dc:subject:adulthood => local:subject:età adulta
dc:subject:advertising => local:subject:pubblicità
dc:subject:aeronautics => local:subject:scienze tecnologiche
dc:subject:aesthetic education => local:subject:educazione estetica
dc:subject:affective development => local:subject:colloquio affettivo
dc:subject:Africa => local:subject:Africa
dc:subject:aggressiveness => local:subject:aggressività
dc:subject:agriculture => local:subject:agricoltura
dc:subject:agronomy => local:subject:agronomia
dc:subject:AIDS => local:subject:malaria
dc:subject:air => local:subject:aria
dc:subject:airport => local:subject:aeroporto
dc:subject:Albania => local:subject:Albania
dc:subject:Albanian language => local:subject:lingua albanese
dc:subject:alcoholism => local:subject:alcolismo
dc:subject:Algebra => local:subject:algebra
dc:subject:America => local:subject:America
dc:subject:American English language => local:subject:lingua inglese
dc:subject:amputation => local:subject:amputazione
```

Figura 8 - Estratto dal file di *mapping* tra il thesaurus ETB (multilingue) e il thesaurus LGI (bilingue)

Recentemente (2005), il BSI [British Standards Institute]<sup>25</sup>, l'organismo nazionale inglese preposto allo studio, alla pubblicazione e alla diffusione di norme e standard tecnici, ha emesso la norma BS 8419-1 e -2, intitolata *Interoperability between*

<sup>25</sup> <<http://www.bsi-global.com/index.xalter>>.

*metadata systems used for learning, education and training. Code for practice for the development of application profiles* (lett.: *Interoperabilità tra sistemi di metadattazione utilizzati nell'apprendimento, nell'istruzione e nella formazione. Manuale per lo sviluppo di profili applicativi*). Nello standard si raccomanda l'uso di XML per la struttura dei metadati e si fa riferimento ai due modelli di metadattazione presentati in questo articolo (il LOM e il modello Dublin Core).

I modelli di metadattazione sono numerosi ed esistono molti schemi specifici, come ad esempio il modello europeo mEDRA<sup>26</sup>, nato in ambito editoriale per la descrizione di prodotti monografici e articoli in pubblicazioni seriali, o il modello americano aap<sup>27</sup>, a cura dell'Associazione degli editori americani (Association of American Publishers, AAP), messo a punto per la descrizione di *e-books*. La prestigiosa Library of Congress presenta, tra gli standard<sup>28</sup> adottati, il MODS [Metadata Object Description Schema]<sup>29</sup>, anch'esso descritto in XML, interoperabile con lo standard MARC<sup>30</sup>, e quindi prevalentemente indicato per applicazioni in ambito biblioteconomico e documentario.

Infine, in contesto educativo, ricordiamo il modello *metadata* messo a punto da EdNA [Education Network Australia]<sup>31</sup>, il *network* australiano di risorse educative, che integra il modello Dublin Core con alcune specifiche educative. Ad esempio, EdNA prevede gli elementi descrittivi "destinatario", "settore", "livello dell'utente" per garantire un migliore utilizzo della risorsa. Un'interessante integrazione della scheda Metadata in EdNA è la parte dedicata alla selezione/valutazione della risorsa e al suo ingresso all'interno della banca dati nazionale, espressa in diversi campi (per es. EdNA.Approver; EdNA.Review; EdNA.Reviewer; ecc.).

## Conclusioni

L'articolo non pretende di essere esaustivo, ma vuole introdurre al tema della metadattazione e dell'interoperabilità *online* delle risorse. Si è cercato di offrire una panoramica internazionale delle varie problematiche, soluzioni e standard che la comunità degli architetti dell'informazione *online* sta trattando. Si rimanda ai riferimenti in bibliografia/sitografia per approfondimenti.

<sup>26</sup> <<http://www.medra.org/en/metadata.htm#faq2>>.

<sup>27</sup> <<http://www.publishers.org/digital/metadata.pdf>>.

<sup>28</sup> <<http://www.loc.gov/standards/>>.

<sup>29</sup> <<http://www.loc.gov/standards/mods/>>.

<sup>30</sup> <<http://www.loc.gov/marc/>>.

<sup>31</sup> <<http://www.edna.edu.au/metadata/>>.

## Bibliografia/Sitografia

- British Standard Institute, *BS 8419 (1-2): Interoperability between metadata systems used for learning, education and training. Code of practice for the development of interoperability between application profiles*. London : BSI Business Information, 2005
- CINECA [Consorzio Interuniversitario per il Calcolo Automatico dell'Italia Nord Orientale], ELEONET European Learning Object Network Project of common interest in the field of trans-european telecommunications networks, in CINECA, *Progetti*. <[http://www.cineca.it/gai/pagina-eleonet\\_progetto.htm](http://www.cineca.it/gai/pagina-eleonet_progetto.htm)> [consultato il 2006-05-22]
- Dublin Core Metadata Initiative, *DCMI Metadata Terms*, 13/06/05, in Dublin Core Metadata Initiative, *Dcml-terms*. <<http://dublincore.org/documents/dcml-terms/>> [consultato il 2006-05-22]
- Education Network Australia (EdNA), *EdNA Metadata Standard v1.1*, in EdNA online, *Metadata*. <<http://www.edna.edu.au/metadata>> [consultato il 2006-05-22]
- ETB partners, *ETB datamodel: minimum set of quality descriptors*, 25/08/03, in European Schoolnet, ETB. <<http://www.eun.org/eun.org2/eun/en/etb/content.cfm?lang=en&cov=5874>> [consultato il 2006-05-22]
- Fini A. - Vanni L., *Learning objects e metadati: quando, come e perché avvalersene*. "I quaderni di Form@re", V. 2, Trento : Erickson, 2004
- IMS Global Learning Consortium, *IMS vocabulary definition exchange best practice and implementation guide* (version 1.0). 2004 <[http://www.imsglobal.org/vdex/vdexv1p0/imsvdex\\_bestv1p0.html](http://www.imsglobal.org/vdex/vdexv1p0/imsvdex_bestv1p0.html)> [consultato il 2006-05-22]
- Institute of Electrical and Electronics Engineers, Inc. (IEEE), *Draft standard for learning object metadata*, 2002, in Institute of Electrical and Electronics Engineers, Inc. (IEEE), *WG12: related materials*. <[http://lts.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://lts.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)> [consultato il 2006-05-22]
- Internet Systems Consortium, Inc. (ISC), *ISC Internet domain survey*, in Internet Systems Consortium, Inc. (ISC), *Internet systems consortium*. <<http://www.isc.org/index.pl>> [consultato il 2006-05-22]
- ISO/IEC JTC1 Information Technology for Learning, Education and Training, *International LOM survey* [N0871], 2004-09-08, in ISO/IEC JTC1 SC36, *Document library*. <<http://jtc1sc36.org/doc/36N0871.pdf>> [consultato il 2006-05-22]
- Library of Congress Network Development - MARC Standards Office, *MODS: Metadata Object Description Schema*, in Library of Congress, *Standards*. <<http://www.loc.gov/standards/mods/>> [consultato il 2006-05-22]
- Lupi M., *Quanto è profondo il web*, [s.d.], in Ad Maiora, *MotoriDiRicerca.IT*. <<http://www.motoridiricerca.it/deepweb.htm>> [consultato il 2006-05-22]

- Panzavolta S., *Learning object, oggetti didattici per l'e-learning*, in Indire [Istituto Nazionale di Documentazione per l'Innovazione e la Ricerca Educativa], *IR: Innovazione e Ricerca*. 2003-01-01 <<http://www.indire.it/content/index.php?action=read&id=56>> [consultato il 2006-05-22]
- Search engine watch*, a cura di D. Sullivan e C. Sherman. ©2005 <<http://searchenginewatch.com/>> [consultato il 2006-05-22]
- Trovabile.org: organizzare l'informazione per renderla (ri)trovabile*, a cura di L. Rosati e C. Gnoli. © 2003-2005 <<http://trovabile.org/>> [consultato il 2006-05-22].



