

# Metadazione e interoperabilità

di Silvia Panzavolta

I **metadati** (o metadata), ossia dati che descrivono altri dati, sono recentemente oggetto di grande attenzione. Dal 1991, anno in cui Tim Berners Lee 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, in cui vengono riportati i **bilioni di pagine indicizzate** dai motori di ricerca più frequentati (Google, Yahoo ecc.).

| Motore di ricerca | Grandezza   | Profondità pagina |
|-------------------|-------------|-------------------|
| Google            | 8,1 bilioni | 101K              |
| MSN               | 5,0 bilioni | 150K              |
| Yahoo             | 4,2 bilioni | 500K              |
| Ask Jeeves        | 2,5 bilioni | 101K+             |

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

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.

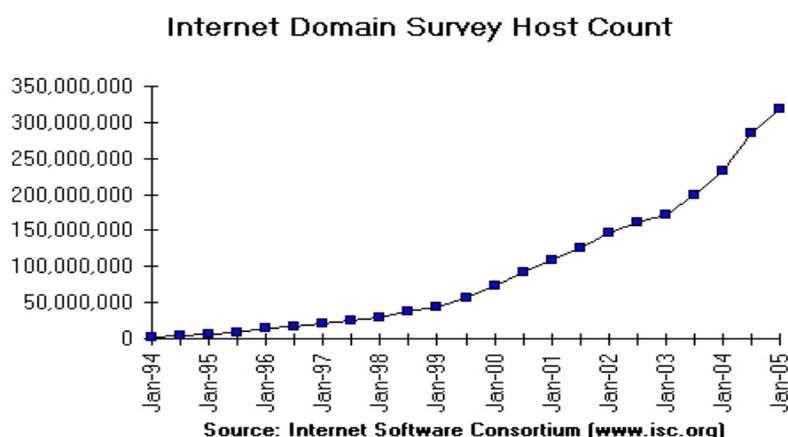


Figura 2 - Grafico relativo all'andamento dei domini Internet dal 1994 ad oggi.

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.

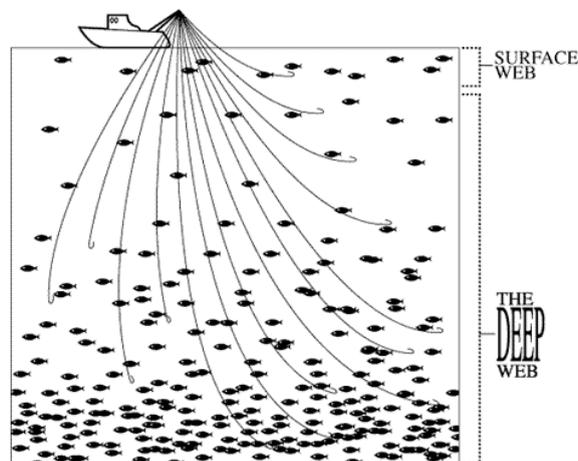


Figura 3 - Rapporto tra Deep Web e Surface Web (fonte: MotoriDiRicerca.IT)

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 per la creazione di un

**catalogo unico delle biblioteche italiane**, cui aderiscono 2382 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 (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 (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 on line**, 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). 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), **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 **DCMI (Dublin Core Metadata Initiative)** che, nel 1999, ha rilasciato la **versione 1.1 del Set di elementi di metadatozione 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 espresso con AAAA-MM-GG (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 (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* (es. banca dati); *evento* (conferenza, seminario, esibizione); *immagine*; *risorsa interattive* (include: applet, learning object multimediali; chat; realtà virtuali); *immagine in movimento* (es. film, animazione, video ecc.); *oggetto fisico*, ossia un'entità non animata e tridimensionale (es. una piramide, una statua); *servizio*, ossia qualsiasi tipo di servizio pensato per un utente finale (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, 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 (<http://www.iana.org/assignments/media-types/>) messo a punto da IANA (Internet Assigned Numbers Authority);

10) **Identificatore** (Identifier): un riferimento non ambiguo alla risorsa nell'ambito di un dato contesto (es. l'URI della risorsa oppure un numero standard, come l'ISBN 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 (es. Firenze <[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)>

, Medioevo ecc.). Anche per questo campo si raccomanda l'uso di un vocabolario controllato, nella fattispecie il Getty Thesaurus of Geographic Names Online

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

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**

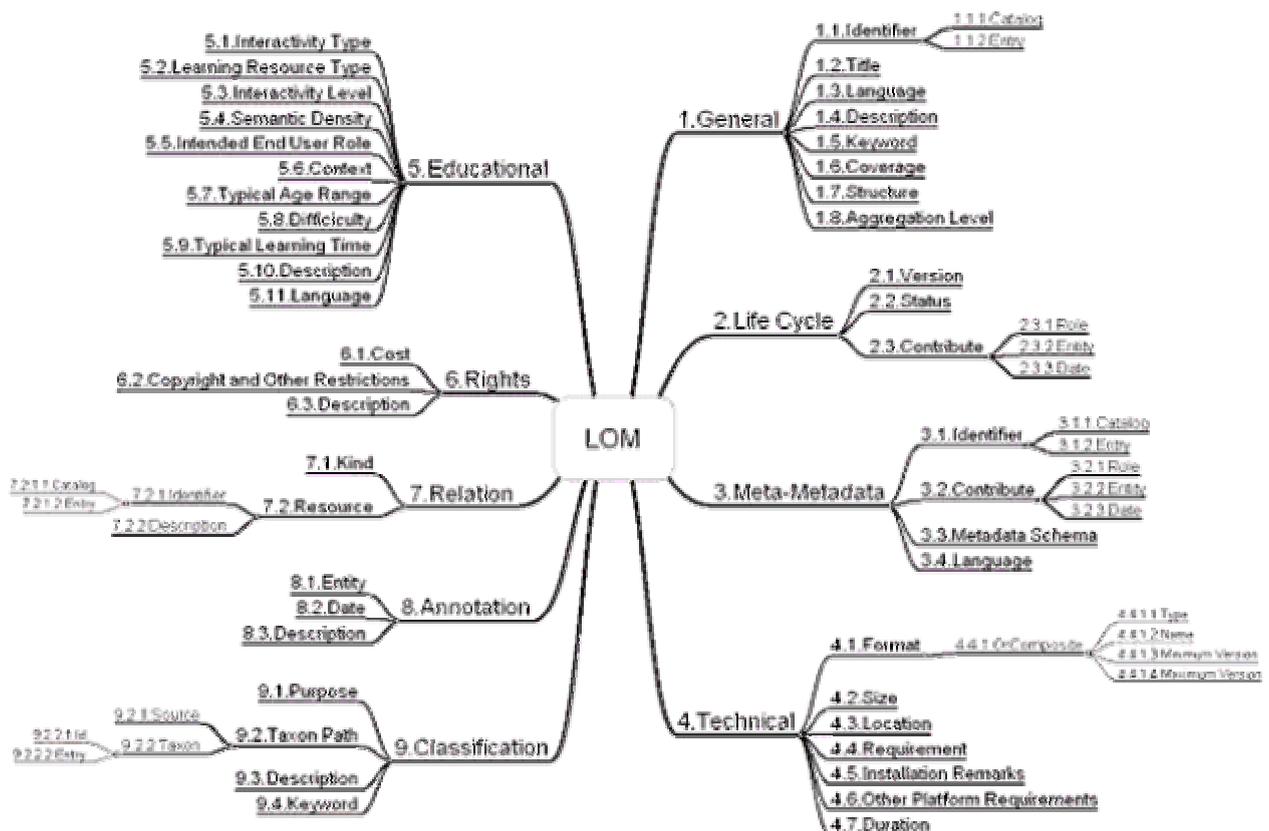
**risorsa se ne può fare**, in particolare **in ambito educativo**. Alcuni elementi sono stati successivamente aggiunti, proprio per sopperire a questa mancanza. La DCMI <<http://dublincore.org/documents/dcmi-terms/>> prevede, infatti, l'aggiunta di elementi come il **livello educativo** del destinatario (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 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** (es. risorsa su attività curricolare; risorsa su attività extracurricolare; 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**  
<<http://www.en.eun.org/etb/survey/d4.2.pdf>>

Lo **schema che intende descrivere risorse educative**, ed in particolare i **Learning Objects**, è il **LOM** (Learning Object Metadata (versione 1.3, 2002) a cura di IEEE (Institute of Electrical and Electronics Engineers) <[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)> (figura

4).



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

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** (9 campi): fornisce informazioni sullo schema Metadata adottato, l'autore/gli autori dello standard, la lingua dello schema (che può essere ? 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 **livello 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* al *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 metatati (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

**N.B.** Come si può osservare, il modello LOM è un modello di metadata **interattivo**. Mentre l'autore o il documentalista (ma il dibattito su chi debba compilare i metadati è aperto!) inserisce le informazioni nella altre categorie, questa è riservata a utilizzatori, valutarori, 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) (<[http://www.eurydice.org/TeeForm/FrameSet\\_EN.htm](http://www.eurydice.org/TeeForm/FrameSet_EN.htm)>).

Il **modello SCORM** (Sharable Content Object reference Model, <<http://www.adlnet.org/scorm/index.cfm>>), 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, <<http://www.adlnet.org/>>), mira prevalentemente ad un'**interoperabilità tecnica dei LO** tra diverse piattaforme di apprendimento (LMS, Learning Management Systems e LMCS, Learning Content Management System). Nelle specifiche SCORM, lo schema Metadata adottato per la descrizione dei LO è il LOM.

Il progetto europeo **CELEBRATE** (Context eLearning with Broadband Technologies)

(<[http://celebrate.eun.org/eun.org2/eun/en/index\\_celebrate.cfm](http://celebrate.eun.org/eun.org2/eun/en/index_celebrate.cfm)>iniziato>, inizio 2002-fine 2004), coordinato da European Schoolnet (<<http://www.eun.org/portal/index.htm>>) 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** (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 (<<http://www.imsproject.org/>>) 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

Poiché pare superfluo illustrare lo schema metadata CELEBRATE, avendo già ampiamente parlato del LOM, si fornisce la versione italiana dello stesso, tradotto a cura di Simona Baggiani, Valeria Biggi, Isabel de Maurissens, Franca Pampaloni e Silvia Panzavolta (<[http://www.indire.it/db/docsrv/Doc/modello%20metadata%20CELEBRATE\\_I.T.doc](http://www.indire.it/db/docsrv/Doc/modello%20metadata%20CELEBRATE_I.T.doc)>).

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 <<http://jtc1sc36.org/doc/36N0871.pdf>>, si consulti il report a cura del gruppo di lavoro SC36/WG4 <<http://jtc1sc36.org/>>, 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 (es. 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 (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 corrispondono.

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 e l'altra il Thesaurus Europeo dell'Educazione. 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 aumenta con l'aumentare 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.

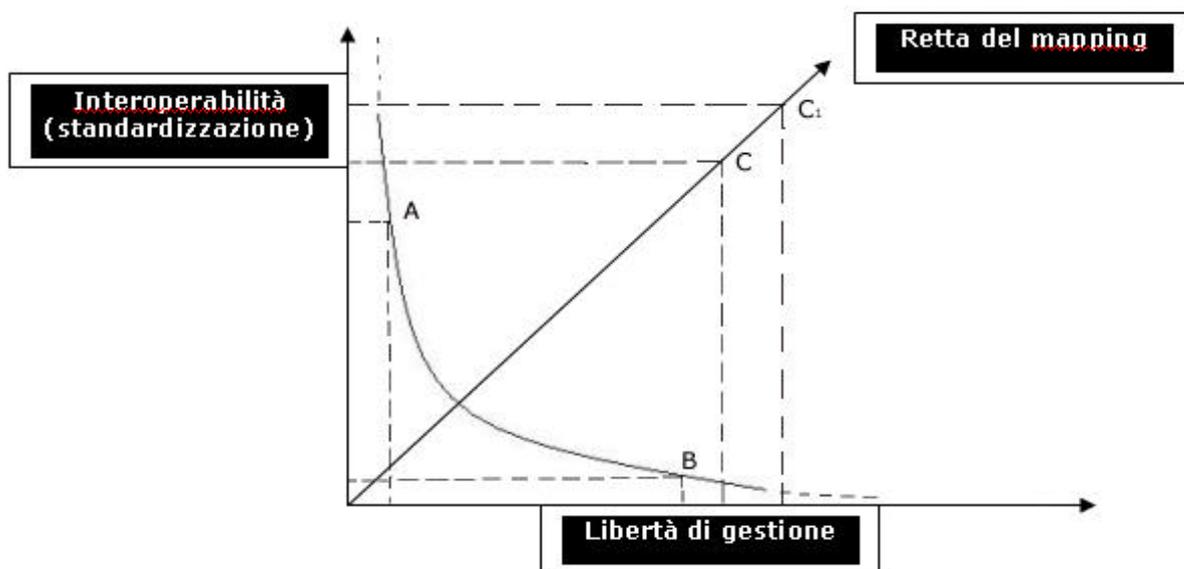
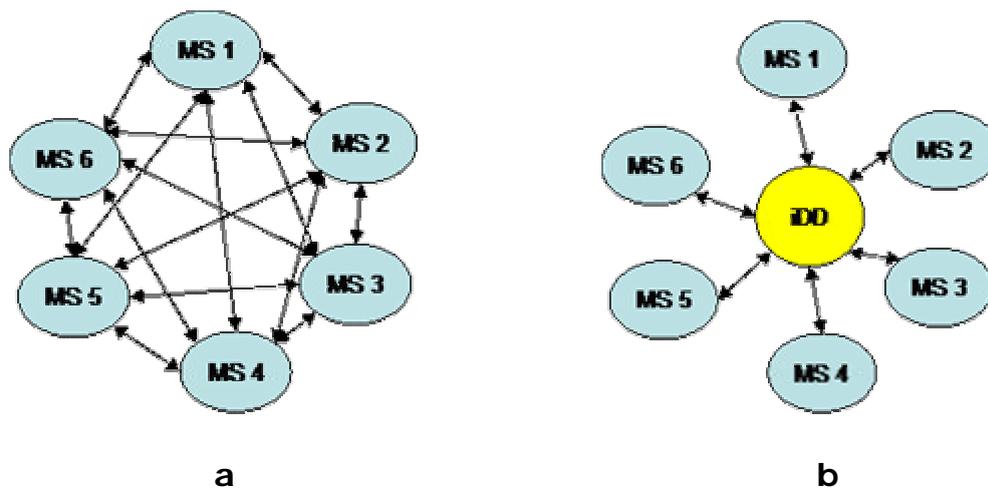


Figura 6 – Rapporto tra standardizzazione e libertà di gestione  
(Autore: Silvia Panzavolta)

In figura viene rappresentato il rapporto tra **standardizzazione** (interoperabilità) e **libertà di creazione**. 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 C<sub>1</sub>).

Il **progetto Eleonet** (European LEarning Object NETwork), cui Indire partecipa in qualità di partner, intende sviluppare un modello metadata per i Learning

Object che non limiti la libertà degli sviluppatori, amministratori e architetti 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**, [http://www.doi.org/handbook\\_2000/appendix\\_4.html#A4.6](http://www.doi.org/handbook_2000/appendix_4.html#A4.6), 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), 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, il sistema di metadatazione 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 Association), tenutasi a Glasgow (Scozia) del 2002, intitolata "**Libraries for Life: Democracy, Diversity, Delivery**" (<http://www.ifla.org/IV/ifla68/>), 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" (<http://www.ifla.org/IV/ifla68/papers/008-122e.pdf>). Il documento contiene una rassegna dei lavori di mappatura tra vocabolari diversi, talvolta anche in lingue diverse, effettuate da varie università, biblioteche e centri di documentazione e ricerca. Fino ad ora, sono stati realizzati mapping **tra vocabolari controllati** nella **stessa lingua** (es. tra il *Library of Congress subject headings, LCSH* e il *Medical subject headings*) e in **lingue diverse** (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** (es. tra il *Library of Congress subject headings*, *LCSH* e la Classificazione Decimale Dewey, *CDD*) e tra **sistemi di classificazione diversi** (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:15th 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:20th century == local:subject:Novecento
dc:subject:21st century -> local:subject:storia contemporanea
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:sviluppo affettivo
dc:subject:Africa == local:subject:Africa
dc:subject:aggressiveness == local:subject:aggressività
dc:subject:agriculture == local:subject:agricoltura
dc:subject:agronomy == local:subject:agrarìa
dc:subject:AIDS -> local:subject:malattia
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:amphibian == local:subject:anfibi
```

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

Recentemente (2005), il **BSI** (British Standards Institute, <<http://www.bsi-global.com/index.xalter>>), 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 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** (<<http://www.medra.org/en/metadata.htm#faq2>>), nato in **ambito editoriale** per la descrizione di prodotti monografici e articoli in pubblicazioni seriali, o il **modello americano aap**, a cura dell'**Associazione degli editori americani** (Association of American Publishers, AAP), messo a punto per la

**descrizione di e-books** (<[www.publishers.org/digital/metadata.pdf](http://www.publishers.org/digital/metadata.pdf)>). La prestigiosa **Library of Congress** presenta, tra gli standard (<<http://www.loc.gov/standards/>>) adottati, il **MODS** (Metadata Object Description Schema, <http://www.loc.gov/standards/mods/>), anch'esso descritto in XML, interoperabile con lo **standard MARC** (<<http://www.loc.gov/marc/>>), 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, <<http://www.edna.edu.au/metadata/>>), 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 (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.

Per eventuali chiarimenti, commenti ecc. si prega di contattare l'autrice all'indirizzo di posta elettronica: [s.panzavolta@indire.it](mailto:s.panzavolta@indire.it)

~~~~~

## 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, BSI Business Information, London, 2005
- ✎ CINECA, 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 11/08/05
- ✎ DUBLIN CORE METADATA INIZIATIVE, DCMI Metadata Terms, 13/06/05, in DUBLIN CORE METADATA INIZIATIVE, *Dcmi-terms*, <<http://dublincore.org/documents/dcmi-terms/>>, consultato il 08/08/05
- ✎ EDUCATION NETWORK AUSTRALIA (EDNA), EdNA Metadata Standard v1.1, in EDNA ONLINE, *Metadata*, <<http://www.edna.edu.au/metadata>>, consultato il 17/08/05

- ✗ 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&ov=5874>>, consultato il 18/08/05
- ✗ FINI A., VANNI L., Learning Objects e metadati: quando, come e perché avvalersene, in *I quaderni di form@re*, v.2, Erikson, Trento, 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 10/08/05
- ✗ INSTITUTE OF ELECTRICAL AND ELECTRONICS (IEEE), Draft Standard for Learning Object Metadata, 2002, in INSTITUTE OF ELECTRICAL AND ELECTRONICS (IEEE), *WG12: Related Materials*, <[http://ltsi.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsi.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)>, consultato il 08/08/05
- ✗ INTERNET SYSTEMS CONSORTIUM (ISC), ISC Internet Domain Survey, in INTERNET SYSTEMS CONSORTIUM (ISC), *Internet systems consortium*, <<http://www.isc.org/index.pl>>, consultato il 05/08/05
- ✗ ISO/IEC JTC1 INFORMATION TECHNOLOGY FOR LEARNING, EDUCATION AND TRAINING, International LOM survey [N0871], 08/09/04, in ISO/IEC JTC1 SC36, *Document Library*, <<http://jtc1sc36.org/doc/36N0871.pdf>>, consultato il 17/08/05
- ✗ LIBRARY OF CONGRESS NETWORK DEVELOPMENT, MARC STANDARDS OFFICE, MODS: Metadata object description scheme, in LIBRARY OF CONGRESS, *Standards*, <<http://www.loc.gov/standards/mods/>>, consultato il 12/08/05
- ✗ LUPI M., Quanto è profondo il web, [s.d.], in AD MAIORA, *MotoriDiRicerca.IT*, <<http://www.motoridiricerca.it/deepweb.htm>>, consultato il 9/08/05
- ✗ PANZAVOLTA S., Learning Object: oggetti didattici per l'e-learning, in INDIRE, *IR: Innovazione e Ricerca*, 01/01/03, <<http://www.indire.it/content/index.php?action=read&id=56>>, consultato il 9/08/05
- ✗ Search Engine Watch, Sullivan D., Sherman C. (a cura di), © 2005, <<http://searchenginewatch.com/>>, consultato il 05/08/05
- ✗ Trovabile.org: organizzare l'informazione per renderla (ri)trovabile, Rosati L., Gnoli C. (a cura di), © 2003-2005, <<http://trovabile.org/>>, consultato il 08/08/05
- ✗ ROSATI L., Architettura dell'informazione, © 2002-2005, <<http://www.lucarosati.it/>>, consultato il 17/08/05