

Modelli per i metadati

Simona Gatta

Garantire l'interoperabilità tra differenti set di metadati è operazione indispensabile per assicurare l'accesso alle risorse in rete. Necessità che si accentua quando ci si trova in presenza di risorse che richiedono descrizioni più complesse di quelle che può assicurare un set semplice come il Dublin Core. Per quanto flessibile e capace di arricchire la sua semantica di base attraverso il meccanismo della qualificazione, il Dublin Core è nato per altri scopi, che egregiamente svolge, ponendosi, come con espressione efficace è stato definito, come una sorta di linguaggio "pidgin" in grado di garantire, di fronte alle dimensioni e alla mutevolezza del Web e alla necessità di assicurarne un minimo ordinamento, un significativo miglioramento nella qualità dell'indicizzazione automatica a bassi costi: cercare di complicarne la struttura significa solo stravolgerne gli assunti di base senza risolvere il problema di arricchire la descrizione di certe risorse, come ha sostenuto in un recente articolo Carl Lagoze.

Invece, fra gli strumenti in grado di migliorare il grado di interoperabilità fra i metadati, alcuni interessanti sviluppi sono offerti, come ci è stato illustrato nell'intervento di Caterina Fasella, da alcuni modelli, sulla scia di quanto proposto dall'IFLA nei *Functional requirements for bibliographic record*, un approccio completamente diverso alle risorse informative e alla loro descrizione catalografica che ha suscitato grande interesse anche in ambienti diversi da quello dei bibliotecari.

Il concetto di modularità espresso nella Warwick Framework, un'architettura al cui interno i diversi pacchetti informativi costituiti dai set descrittivi specifici, creati e aggiornati dalle singole comunità di esperti, vengono a integrarsi, e che è alla base della Resource Description Framework, si trova a dover fronteggiare la sfida di garantire l'interoperabilità fra tali pacchetti.

Diversi i meccanismi utilizzati, in primo luogo naturalmente le mappature. Che però, per essere realmente efficienti e garantire un elevato grado di qualità dovrebbero essere del tipo uno a uno; come abbiamo potuto constatare anche nel nostro lavoro, quando la mappatura riguarda più set contemporaneamente le equivalenze non sono sempre facili da costruire, e non sempre sono convincenti. Questo naturalmente pone un problema di dimensioni: è impossibile pensare di poter fronteggiare gli innumerevoli set predisposti dalle comunità di esperti, le loro modifiche e aggiornamenti, i nuovi set che potrebbero essere creati, con questo unico strumento.

Un tentativo interessante è stato quello dell'UKOLN che nell'ambito del progetto DESIRE ha proposto la costruzione di un registro di metadati usando per le mappature quello che viene definito un "sostrato semantico": i diversi set non vengono mappati direttamente fra di loro, ma usando questa sorta di linguaggio di traduzione come intermediario.

Un modello logico si propone appunto di essere una sorta di interlingua, uno strumento che consenta di passare agevolmente da una semantica a un'altra, senza che ciò comporti perdita di significato o ambiguità.

In particolare, sembra attagliarsi bene alla natura delle risorse in rete e alle esigenze di interoperabilità fra le diverse semantiche che le descrivono, un modello del tipo event-awareness, ovvero che pone al centro della propria costruzione gli eventi che scandiscono le tappe dell'esistenza di una risorsa.

Già era implicita nella definizione delle entità del primo gruppo di FRBR, i prodotti dell'attività intellettuale o artistica, il concetto di evoluzione temporale: una risorsa nasce come creazione intellettuale, entità astratta che prende forma in una o più realizzazioni intellettuali; queste espressioni di un'opera a loro volta prenderanno corpo in una o più concretizzazioni fisiche, che potranno essere disseminate in uno o più esemplari. Ciascuno di questi passaggi può essere reso esplicito caratterizzando nella maniera adeguata l'avvenimento che ne è all'origine e che comporta una trasformazione della risorsa e, di conseguenza, degli attributi che la caratterizzano. Esplicitare gli avvenimenti consente anche l'individuazione più agevole di chi vi ha avuto parte e con quale

ruolo, nonché del contesto in cui si svolgono. Sono questi concetti che aiutano a comprendere meglio le risorse e la loro evoluzione e quindi a descriverle. E sono concetti comuni: un tavolo, un documento, una seduta, un libro sono *creazioni*, fatte e usate da *agenti*, creazione e uso che significano *eventi* e *azioni* che si svolgono in *luoghi* e *date* determinati.

Porre al centro dell'analisi gli avvenimenti che scandiscono l'esistenza di una risorsa significa fissarne una sequenza di immagini, di istantanee potremmo definirle, in un determinato luogo e a una certa data, che mantengono alcune caratteristiche della risorsa, da un passaggio all'altro, ed altre mutano. Gli eventi nell'arco di vita di una risorsa sono potenzialmente infiniti: stabilire quali fra essi, e dunque quali caratteristiche nuove della risorsa sono di interesse per una determinata comunità, spetterà a quest'ultima, che utilizzerà per descriverla lo standard ritenuto più appropriato (ad esempio, di una seduta parlamentare, che avrà sempre lo stesso numero distintivo, la stessa data e avverrà nello stesso luogo, i servizi parlamentari potrebbero essere interessati agli eventi che incidono sulla tecnica procedurale, mentre uno studio legale potrebbe essere piuttosto interessato al contenuto della discussione su un determinato progetto di legge; di un'opera d'arte, che avrà sempre la medesima attribuzione di autore e di epoca, posso voler ricostruire la storia dei passaggi di proprietà a scopo fiscale oppure per ricostruire la storia degli interventi di restauro: infiniti gli esempi possibili). Ogni comunità dunque sarà distinta dagli eventi che hanno significato all'interno di essa e che sceglierà di descrivere, tralasciandone altri. Anche il livello di granularità al quale decidere di arrivare dipenderà dalle singole comunità: se scegliere di analizzare tutti gli eventi restaurativi o limitarsi a quelli che hanno avuto un forte impatto sull'opera, se prendere in considerazione solo l'approvazione che trasforma un progetto di legge in legge, o ricostruirne tutto l'iter legislativo.

L'importanza riconosciuta da varie comunità, come quella delle biblioteche, degli archivi, dei musei, dell'e-commerce e della gestione dei diritti di proprietà intellettuale, alla centralità dei processi e degli eventi nella descrizione delle risorse, e che si è concretizzata nella costruzione di modelli come FRBR, CIDOC/CRM, INDECS, ha indotto a tentare un approccio più generale, come quello proposto nel modello ABC dal progetto Harmony, iniziato nel 1999, con l'intento di investigarne l'applicabilità, in particolare, nelle biblioteche digitali multimediali.

L'attenzione del progetto si è focalizzata sull'analisi delle entità, dei loro attributi e delle relazioni che sono comuni nei diversi set di metadati e sui differenti requisiti descrittivi propri delle singole comunità, sulla possibilità di determinare un set di base di entità e proprietà comuni che possano costituire i punti di aggancio per i singoli vocabolari descrittivi, e sulla costruzione di un'architettura concettuale in grado di estrapolare le entità comuni dai singoli vocabolari e utilizzarle per rendere più semplici le mappature e le integrazioni fra set di metadati. Risultato di questo lavoro sono una serie di principi e di assunti di fondo (come, ad esempio: l'universo ABC è costituito da risorse univocamente identificabili; da cui, la necessità di identificatori unici per tutte le risorse; le proprietà, o attributi, esprimono relazioni fra risorse; a molteplici punti di vista corrispondono descrizioni di diversa complessità), che si traducono in un vocabolario che definisce e dichiara un set di base di classi (risorse, eventi, agenti, tempo, luogo), in alcune regole di gerarchia per le mappature, regole per consentire la trasformazione di una descrizione più ricca e articolata in una più semplice e il passaggio inverso, regole che consentano di passare da una descrizione basata sull'evento ad una che abbia al suo centro la risorsa.

[Lucido: meccanismo di trasformazione da evento a documento a standard; da art. di Jane Hunter]

Molte le questioni ancora aperte, naturalmente, dalla mancanza di un metodo soddisfacente per la rappresentazione grafica del modello che sia in grado di descrivere sia i documenti che gli eventi e le loro relazioni, a come gestire i valori degli attributi che talune risorse possono ereditare da altre, come assicurare una corretta e non ambigua trasformazione da una struttura complessa ad una più semplice quando ci si trova in presenza di più agenti che operano all'interno di un evento e questo dà origine a più di una risorsa, o quando più di una risorsa va a confluire in un unico evento, quali i meccanismi e gli strumenti di formalizzazione più appropriati per renderlo applicabile (un

tentativo di utilizzare XML Schema e XSLT per mappare sul modello ABC alcuni set di metadati come il Dublin Core, MPEG-7, ID3 ha dimostrato la limitazione di questi strumenti nelle mappature se non supportati da strumenti più potenti, come un metatesauro, in grado di fornire la conoscenza semantica necessaria).

Altri test sono in corso per verificare se quello che sulla carta sembra funzionare, ed essere molto promettente, non incontri poi ostacoli insormontabili nella realtà. Ad esempio, è in corso una collaborazione fra Harmony e il CIMI, iniziata nella seconda metà dell'anno scorso e il cui rapporto finale è previsto per questo aprile, per testare il modello ABC su un certo numero di record messi a disposizione da alcune istituzioni museali e l'uso di strumenti come XML Schema, XSLT, RDF per l'implementazione: nella descrizione dell'iniziativa si avverte come si tratti di un progetto puramente di ricerca perché "the ideas may not work out".

Un altro interessante sviluppo è offerto dalla collaborazione con DELOS allo scopo di verificare se è possibile la convergenza fra il modello ABC e il CIDOC/CRM nel quadro di una più generale armonizzazione fra i metadati (un primo incontro si è tenuto qui a Roma nei giorni scorsi, 26-27 marzo). Il Conceptual Reference Model è un modello sviluppato nella comunità dei musei a partire dal 1996 e pubblicato nella sua prima versione ufficiale nel 1998. Si rivolge in particolare quindi agli oggetti museali, ma possiede le potenzialità per dialogare fruttuosamente anche con le biblioteche e gli archivi. È un modello sia object-oriented che event-awareness, stabile nella struttura ma estensibile ed è stato proposto come norma ISO. Una serie di test cui è stato sottoposto per verificarne l'utilità e la completezza, ne hanno dimostrato la capacità di proporsi come mediatore fra set diversi di standard: in particolare, sono state create delle mappature verso AMICO, SPECTRUM, Dublin Core qualificato, EAD. Sulla base di questa esperienza e della discussione che ne è scaturita nel corso di un incontro tenutosi nel giugno 2000, sono state proposte una serie di estensioni al modello che ne hanno ampliato e meglio definito una serie di concetti e relazioni, rafforzandone quindi il ruolo di strumento di mediazione nei confronti degli altri standard. Il CRM si definisce una ontologia per l'informazione sul patrimonio culturale, ovvero cerca di descrivere in un linguaggio formale i concetti impliciti ed espliciti e le relazioni inerenti la documentazione del patrimonio culturale. Un'ontologia di questo tipo ha bisogno, per non cadere nell'arbitrarietà, di criteri chiari ed espliciti, che sono stati stabiliti dagli esperti del CIDOC esaminando una serie di fattori (come il punto di vista degli utenti ai quali si intendeva rivolgersi, le attività che si volevano sostenere, il tipo di oggetti da prendere in considerazione, il grado di dettaglio e precisione necessario per garantire un adeguato livello di qualità, la complessità delle tecnologie coinvolte).

Un'analisi di questo tipo è l'approccio proposto nell'ambito della cooperazione fra DELOS e Harmony per sviluppare i criteri adeguati a inserire ABC e CRM in uno scenario più ampio che dovrebbe condurre alla creazione di una superontologia, attraverso la fusione dei due modelli o un loro sviluppo coordinato, allo scopo di raggiungere un sempre maggior grado di armonizzazione fra i metadati.