



Linked Heritage: una piattaforma cooperativa di gestione della terminologia per una rete di tesauri multilingue e di vocabolari controllati

Marie-Veronique Leroi

Il contesto

Il web semantico e i linked data sono attualmente principi ben noti del web. L'applicazione delle regole dei linked data è stata considerata una priorità per lo sviluppo di Europeana, la Biblioteca Digitale Europea. Sebbene la normalizzazione dei metadati sia stata oggetto di studio per anni, l'importanza delle terminologie per comprendere e sfruttare questi metadati in modo strutturato è venuta alla ribalta grazie ai linked data e al web semantico.

Tim Berners-Lee definisce il web semantico «the Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it». Oggetto del web non sono più i documenti ma i dati. Quindi si può parlare di linked data in questi termini «The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data». Il web dei dati ha un senso solo

se i dati sono collegati tra di loro. Molti progetti europei hanno contribuito a Europeana con contenuti specifici. Il patrimonio culturale è ricco grazie alla sua diversità: gallerie, biblioteche, archivi e musei, noti anche con l'acronimo GLAM, hanno una percezione diversa dell'oggetto culturale. Oltre alla diversità dell'oggetto in sé, c'è una grande differenza nei linguaggi e nelle terminologie specifiche. È necessaria quindi un'armonizzazione delle terminologie a livello europeo per consentire una migliore comprensione dei contenuti disponibili in Europa. Il progetto Michael¹ è stato uno dei primi progetti europei a prendere in considerazione la questione della pluralità delle lingue in ambito europeo. Di fatto, questo portale offre un accesso multilingue al patrimonio culturale digitale in dodici lingue. L'Associazione Michael Culture creata dopo la fine del progetto per sostenere l'attività del portale europeo, è stata coinvolta nei progetti europei che hanno contribuito a Europeana. Athena è stato uno dei maggiori progetti che ha contribuito con il proprio contenuto a Europeana fornendo circa un milione e ottocentomila oggetti. In questa presentazione viene illustrato come il lavoro sulle terminologie, avviato nell'ambito del progetto Athena, viene ora riusato e implementato nel progetto Linked Heritage. Athena è stato uno dei maggiori progetti che ha contribuito con il proprio contenuto a Europeana fornendo circa un milione e ottocentomila oggetti. In questa presentazione viene illustrato come il lavoro sulle terminologie, avviato nell'ambito del progetto Athena, viene ora riusato e implementato nel progetto Linked Heritage.

Athena

Il progetto Athena, avviato nel novembre 2008 e concluso ad aprile 2011, aveva come obiettivo quello di fornire i contenuti dei musei eu-

¹<http://www.michael-culture.org>.

ropei a Europeaana. L'Associazione Michael Culture è stata coinvolta nel progetto Athena come leader nell'ambito del pacchetto di lavoro dedicato alla terminologia e al multilinguismo (WP4). Gli obiettivi principali di questo pacchetto di lavoro erano quelli di fornire ai musei europei raccomandazioni per la gestione della terminologia. Quindi siamo partiti con la definizione e la presentazione di diverse tipologie di risorse terminologiche. Infatti, poiché molte risorse terminologiche sono definite con l'espressione «vocabolario controllato», è stato elaborato uno schema sintetico per guidare le istituzioni a rispondere alla nostra indagine. Obiettivo dell'indagine, lanciata tra i partner di Athena, era ottenere uno stato dell'arte esaustivo e fornire un quadro preciso circa l'uso e i metodi della terminologia nei musei europei. Tutti i risultati relativi allo stato dell'arte sono stati resi disponibili nel wiki di Athena e di Linked Heritage e nel sito del progetto.²

I risultati dell'indagine ci mostrano che la maggior parte delle istituzioni culturali, il 40% (su un totale di 105 risposte) usa una terminologia di tipo tesaurale. Effettivamente la maggior parte delle istituzioni culturali usa un vocabolario controllato strutturato. Il thesaurus non solo offre relazioni di tipo gerarchico e associativo tra i descrittori, ma è anche uno strumento molto semplice e potente per l'indicizzazione e la catalogazione.

SKOS (Simplified Knowledge Organisation System), riconosciuto come raccomandazione W3C nell'agosto 2009 e sistema di organizzazione della conoscenza richiesto da Europeaana, viene usato sempre più spesso per convertire tutte le risorse terminologiche in un formato interoperabile. Solamente poche tra le istituzioni che hanno risposto all'indagine avevano già una terminologia convertita in formato Simplified Knowledge Organisation System (SKOS). La maggior parte dei musei gestisce i propri vocabolari all'interno

²<http://www.athenaeurope.org/getFile.php?id=398>.

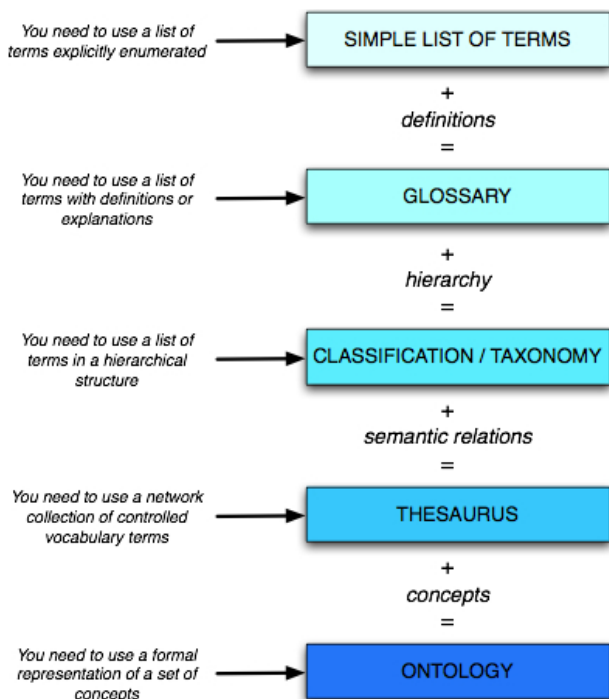


Figura 1: Tipologie di risorse terminologiche

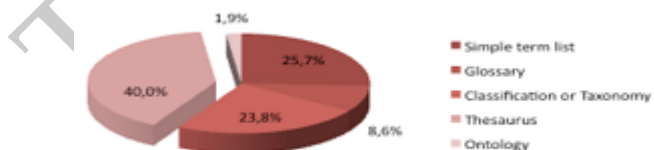


Figura 2: Differenti terminologie

del sistema di gestione delle loro collezioni. Questi strumenti sono generalmente proprietari e spesso consentono la conversione in XML e in alcuni casi in formato SKOS. Oltre all'identificazione delle risorse terminologiche utilizzate dai musei europei, il secondo fine del pacchetto di lavoro era definire alcune linee guida ed un tutorial per SKOS.

Raccomandazioni

La prima fase riguarda l'elaborazione della propria terminologia. Come dire, in questa fase un'istituzione gestisce internamente la propria terminologia in modo da creare un thesaurus in una prospettiva umana. Abbiamo indicato le varie fasi-chiave per l'istituzione che si accinge a creare per proprio conto una nuova terminologia o ad adattare una terminologia già in uso al fine di ottimizzare la descrizione delle proprie risorse digitali per Europeana. Queste operazioni sono prioritarie in quanto determinano le due fasi successive. In questa fase l'istituzione deve considerare il dominio (i domini) coperto (coperti) dalla terminologia, chi sono i potenziali utenti che la utilizzeranno e quali saranno le lingue in cui la terminologia sarà disponibile. I termini e la loro organizzazione all'interno della struttura del thesaurus vengono definiti in questa fase.

La seconda fase consiste nel rendere la terminologia interoperabile. Si tratta ora di estrarre la terminologia dal contesto del museo. Indubbiamente quando una terminologia è convertita in formato SKOS assume l'ottica della macchina. Con questa seconda fase siamo nella prospettiva del web semantico. A questo punto l'istituzione deve valutare se il formato SKOS è quello pertinente per il tipo di terminologia gestita. SKOS, per esempio, non è il formato appropriato nel caso di nomi di autori/persona. Dopo questa valutazione si passa all'obiettivo principale di questa fase, e cioè alla

conversione della terminologia nel formato SKOS. Esistono tool in grado di validare i risultati in questo formato.

Infine, rivolgiamo alle istituzioni culturali le nostre ultime raccomandazioni in quanto queste riguardano il collegamento della loro terminologia con altre. La terza e ultima fase per un'istituzione consiste nell'essere visibile in Europa in una prospettiva di rete, integrando la propria terminologia in una rete di terminologia convertita in formato SKOS. Nell'ambito di questa fase la definizione di alcuni metadati per descrivere i dettagli amministrativi della terminologia è un passaggio necessario. Sulla base dei metadati forniti dalle altre terminologie saranno identificate le risorse terminologiche che possono essere mappate con quella dell'istituzione. Con questa ultima fase entriamo nella prospettiva dei linked data.

Da Athena a Linked Heritage

Il WP3 di Linked Heritage si basa sull'eredità lasciata dal WP4 del progetto Athena ed è stato organizzato per affrontare nel modo migliore i seguenti aspetti: la gestione dei contenuti e gli sviluppi tecnologici relativi alla gestione della terminologia. Nell'ambito del WP4 del progetto Athena, raccogliendo le esperienze dei musei e di altri domini, siamo riusciti a individuare il processo più logico e le esigenze funzionali relativi alla gestione, all'interoperabilità semantica e all'arricchimento delle terminologie. Nel fare ciò nell'ambito del progetto sono stati individuati alcuni casi di applicazione e definito un termine di riferimento. Il lavoro sui casi di applicazione ha reso evidente che in primo luogo era necessaria una specifica del diagramma di flusso di lavoro per una produzione e un'armonizzazione su base collaborativa delle terminologie relative ai patrimoni culturali. Con l'aiuto del gruppo di lavoro esperto sulle terminologie

e dei casi di applicazione definiti precedentemente, si è considerato il flusso di lavoro rappresentato in figura 3.



Figura 3: Riferimento - Flusso di lavoro

Tale flusso di lavoro ha consentito di valutare gli strumenti esistenti, le interfacce e i metodi adeguati per questo lavoro. Sfortunatamente nessuno degli strumenti valutati era in grado di gestire di fatto l'intero processo così come viene mostrato nella figura 3. Alcuni strumenti ci si avvicinavano, offrendo le opzioni di ricerca e navigazione, oltre alla mappatura semantica e all'arricchimento in un ambiente collaborativo. Ma non c'era nessun software completo disponibile che, in primo luogo, offrisse la possibilità per le istituzioni di patrimoni culturali di caricare, registrare e convertire in formato SKOS le terminologie, prima di procedere alle fasi successive. Questo è stato il vero punto debole del flusso di lavoro dei tool esistenti, in quanto la maggior parte delle istituzioni di patrimoni culturali utilizzano proprie terminologie di riferimento prodotte *in-house* e non hanno le risorse disponibili per gestirle in un formato standard interoperabile come SKOS. Questa lacuna tecnologica è stata ampiamente considerata nell'ambito del WP3 di Linked Heritage in quanto verrà sviluppato un software completo per affrontare tutte le fasi e i processi individuati con Athena.

Dalla teoria alla pratica: Terminology Management Platform (TMP)

Terminology Management Platform (TMP) sarà il software completo per la gestione della terminologia e seguirà lo stesso flusso di lavoro strutturale di cui sopra. Linked Heritage va oltre il lavoro svolto nell'ambito di Athena, con l'intento di sviluppare il prototipo di un tool in grado di avere a che fare con le varie fasi del flusso di lavoro definite, e quindi di diminuire gli sforzi economici di cui l'istituzione si deve fare carico nel momento in cui si impegna a condividere la terminologia in un formato scambiabile con la comunità e con Europeana. Il lavoro svolto nel WP4 del progetto Athena ha avuto come esito la risoluzione di progettare e implementare un software integrato per la gestione della terminologia che consenta ad ogni istituzione di gestire la propria terminologia secondo le regole di gestione proprie di Europeana. Nel Linked Heritage questa soluzione diventerà realtà con lo sviluppo di un prototipo di Terminology Management Platform (TMP) per il settore dei patrimoni culturali, al fine di creare in modo collaborativo una rete di terminologie multilingue, interconnesse in un formato Europeana compatibile (SKOS). Nello sviluppo di tale piattaforma saranno fatte confluire le esperienze di quattro partner tecnici, nello sforzo congiunto di creare un ambiente integrato per la gestione della terminologia. L'approccio, basato sulla condivisione delle esperienze e sull'integrazione di tool già esistenti in un unico ambiente web, ci consente di selezionare e combinare le caratteristiche di buone pratiche tecnologiche in modo efficiente sia per quanto riguarda i tempi che i costi. Alla luce dei risultati scaturiti da Athena e i primi obiettivi conseguiti nell'ambito del WP3 di Linked Heritage, sono qui di seguito individuate le caratteristiche per la TMP:

- essere un servizio del web: per un lavoro di collaborazione

online;

- avere una GUI amichevole: adattata per l'uso di non addetti ai lavori di musei, biblioteche e archivi europei;
- mettere insieme componenti open-source: tale servizio deve rimanere indipendente rispetto a codici e formati proprietari;
- essere strutturato in modo logico con un flusso di lavoro intuitivo: l'utente deve poter trovare i percorsi da seguire a seconda delle proprie esigenze;
- essere sufficientemente flessibile per adattarsi a nuovi standard: che succederebbe se SKOS venisse aggiornato in una nuova versione o evolvesse verso una descrizione di tipo ontologico?

Come detto in precedenza, quattro partner tecnici sono coinvolti nello sviluppo di TMP. Ciascuno di essi ha la propria esperienza in termini di sviluppo e usabilità e questa diversità contribuisce ad arricchire l'intero pacchetto del tool. Lo schema in figura 4 nella pagina seguente rappresenta l'architettura tecnica di TMP

Digicult, il partner tecnico tedesco, sta già sviluppando e gestendo un tool di editing, xTree. L'Istituto di Scienza e Tecnologia (IST) portoghese ha già partecipato a progetti che hanno contribuito a Europea ed in particolare allo sviluppo di archivi di metadati. Questa conoscenza ed esperienza viene ora applicata alla TMP di Linked Heritage in quanto Istituto di Scienza e Tecnologia (IST) è responsabile dello sviluppo dell'archivio terminologico di TMP. L'Università Tecnica Nazionale di Atene era già coinvolta nel progetto Athena e ha creato il tool Mint utilizzato per l'ingestion dei metadati. Per poter assistere le istituzioni che utilizzeranno TMP si è deciso di usare lo stesso processo di autenticazione, usato nel tool Mint, per l'ingestion in modo tale che le istituzioni non avranno bisogno

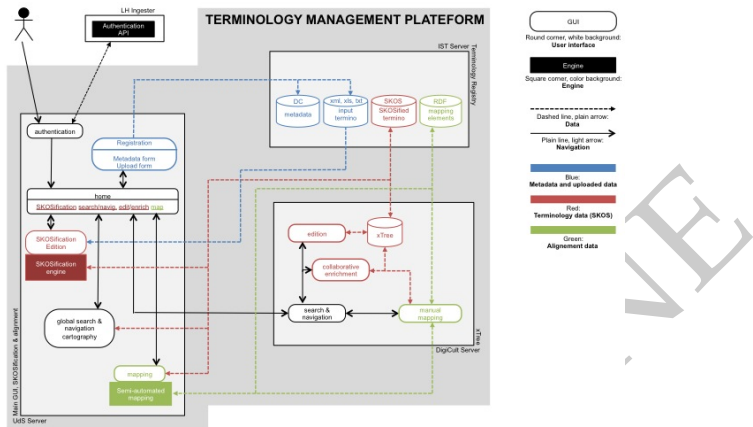


Figura 4: Schema per l'architettura tecnica della TMP

di registrarsi nella TMP se stanno già utilizzando Ingester. Infine, l'Università della Savoia francese, che ha una grande esperienza nell'ingegneria della conoscenza, è responsabile dello sviluppo di un tool per la conversione in formato SKOS e per la navigazione visiva. Ma, più d'ogni altra cosa, l'Università della Savoia si sta impegnando per mettere insieme le diverse tecnologie e i moduli degli altri partner tecnici.

Conclusioni

Linked Heritage eredita dal progetto Athena la possibilità di riutilizzare il tool Mint per l'ingestion e il formato LIDO. Nel campo delle terminologie va persino avanti mettendo in pratica tutte le raccomandazioni e il flusso di lavoro definiti in Athena. Terminology Management Platform significherà molto per le istituzioni che quindi saranno autonome per quanto riguarda la gestione e la map-

patura delle loro terminologie. Tutto ciò renderà inoltre possibile una maggiore consapevolezza degli aspetti relativi al web semantico e ai linked data.

Ai fini di una corretta indicizzazione, si invitano i lettori a citare esclusivamente il testo in lingua inglese; l'unico, infatti, che presenta l'indicazione del numero di pagina, l'abstract, le keywords e le date del processo redazionale.

Leroi, M.V. "Linked Heritage: a collaborative terminology management platform for a network of multilingual thesauri and controlled vocabularies". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #5471. DOI: [10.4403/jlis.it-5471](https://doi.org/10.4403/jlis.it-5471). Web.

