

Linked data per la gestione della conoscenza condivisa nelle biblioteche, negli archivi e nei musei (LAM)

Sintesi del contesto

Biblioteche, archivi e musei (LAM) possiedono grandi quantità di dati e risorse che finora sono spesso rimaste nascoste all'interno dei cataloghi e degli archivi. Sfruttare il potenziale di queste informazioni diffondendole a un pubblico più ampio rappresenta un'opportunità per arricchire ulteriormente il World Wide Web, promuovere una cultura di apertura alla conoscenza e creare una serie di vantaggi per tutti gli elementi nella catena dell'informazione. In questo modo il patrimonio culturale mondiale può essere preservato a beneficio delle generazioni future e la ricerca può essere mantenuta nelle sue lingue originali, conservandone la varietà e la vitalità culturale.

I linked data sono lo strumento che può rendere tutto questo possibile.

Il termine linked data si riferisce a un insieme di regole per la pubblicazione e la connessione di dati strutturati sul Web e si basa sul concetto di relazione qualificata. Il World Wide Web si fonda sulla creazione di collegamenti ipertestuali tra i documenti Web, che ne costituiscono la base per la navigazione. L'idea centrale dei linked data consiste nell'usare collegamenti ipertestuali non solo per identificare i documenti Web, ma anche entità arbitrarie del mondo reale con lo scopo finale di creare una vera e propria rete di dati univoci e certificati. Viene così superato il problema dell'integrazione di cataloghi di oggetti diversi, grazie a un sistema che li interconnette indipendentemente dalla loro tipologia (libri, opere etc.) e dalla loro provenienza (biblioteche, musei, archivi etc.), creando un tessuto di informazioni estremamente più funzionale e completo per chi ricerca e nello stesso tempo rendendo la conoscenza condivisa uno strumento collaborativo per le istituzioni culturali.

I linked data pongono quindi le basi per ulteriori miglioramenti nelle dinamiche di collaborazione tra istituzioni e per il riutilizzo dei dati in contesti diversi, consentendo identificazione e possibilità di discovery più efficienti basate su Web.

Questo nuovo approccio è per biblioteche, archivi e musei un'occasione preziosa per creare legami tra i cataloghi e fornire agli utenti strumenti di ricerca più sofisticati e sempre in linea con l'evoluzione delle nuove tecnologie web, mentre per i produttori di dati si configura come un passo decisivo per la creazione e la promozione di nuove forme di cooperazione.

I potenziali benefici per l'intero settore degli studi umanistici, che ora più che mai rischia la marginalizzazione, sono significativi.

Flessibile e sostenibile, questo approccio cooperativo consente la creazione di un sistema ben strutturato per l'organizzazione dei dati, valorizzando la specificità delle varie tradizioni culturali e sfruttando le opportunità offerte dalle nuove tecnologie. Le figure chiave nel settore potranno contribuire al fianco di organizzazioni più piccole, pronte a condividere la propria esperienza e l'unicità delle proprie risorse.

Per rispettare la nuova filosofia aperta di condivisione e riutilizzo dei dati, anche il concetto di record di autorità tradizionale deve evolversi, spostando il focus della discussione dallo stabilire se il controllo debba avvenire a livello centrale o locale, verso un concetto più flessibile incentrato sull'identificazione dell'entità e delle relazioni tra entità. Biblioteche, archivi, musei, editori e fornitori saranno protagonisti nella generazione di nuovi dati e nella scoperta di nuove risorse, superando i confini dei loro domini specifici per creare opportunità di arricchimento dei dati che prima sarebbero state impensabili.

In un contesto di rinnovamento dei linguaggi e degli standard di catalogazione, che ha visto negli ultimi anni l'emergere delle linee guida RDA (Resource Description and Access) e la pubblicazione del modello IFLA Library Reference Model (LRM), nasce il modello di dati BIBFRAME, che è attualmente oggetto di discussione e sviluppo all'interno della comunità bibliotecaria e non solo. L'intenzione alla base del framework è quella di aprire le possibilità offerte dai linked data a biblioteche, archivi e musei, fornendo maggiore interoperabilità, visibilità e reperibilità per tutti i tipi di risorse.

Lo scopo dei componenti e degli strumenti sviluppati nell'ambito della famiglia di iniziative Share basate sui linked data, che include Share-VDE (Virtual Discovery Environment), Share-ART (con le biblioteche di storia dell'arte dell'Istituto Max Planck) e Share-MUSIC (una sperimentazione nel settore musicale) è quello di creare un ambiente utile per la gestione della conoscenza. Le interfacce di ricerca avanzate miglioreranno l'esperienza d'uso e forniranno risultati più ampi alle biblioteche, agli archivi e ai loro utenti. Bibliotecari, archivisti e curatori di musei trarranno vantaggio da progressi come lo sviluppo continuativo delle funzioni di catalogazione secondo gli standard del web semantico, l'integrazione dei processi di un ambiente collaborativo con sistemi e strumenti locali. L'obiettivo è quello di far sì che una grande mole di dati, che spesso rimangono nascosti o inespressi in un catalogo tradizionale, riveli finalmente la sua ricchezza all'interno delle collezioni esistenti.

Nello svolgersi dell'iniziativa i record di biblioteche, archivi e musei vengono trasformati in linked data, tramite l'identificazione dell'entità, la riconciliazione e l'arricchimento dei dati. Gli attributi sono utilizzati per identificare in modo univoco una persona, un'opera o altre entità, con le forme varianti riconciliate per formare un cluster di dati riferito alla stessa entità. I dati sono successivamente riconciliati con i tradizionali record di autorità e altre fonti esterne per arricchirli e creare una rete di informazioni e risorse. Il risultato è un database di relazioni, aperto all'intera comunità, e una knowledge base di cluster in RDF comune. Il database utilizza i paradigmi del web semantico ma consente anche alle istituzioni partecipanti di continuare a gestire i propri dati in modo indipendente.

Tra le principali aree di attività:

- arricchimento dei record MARC con URIs;
- conversione dei dati in MARC in RDF usando il vocabolario BIBFRAME (e altre ontologie aggiuntive, se necessario);
- creazione di una piattaforma di discovery basata su un adattamento del modello BIBFRAME, sviluppata per fornire un servizio di discovery in linked data;
- creazione di un database di relazioni e cluster di risorse (Sapientia, la Cluster Knowledge Base) accessibile in RDF;
- implementazione di componenti per l'interazione diretta con i dati, che consentono validazione, aggiornamento, controllo a lungo termine e manutenzione dei cluster e degli URI che identificano le entità;
- procedure di aggiornamento automatico o batch dei dati;
- distribuzione in batch/automatizzata di dati alle biblioteche;
- implementazione progressiva di ulteriori casi d'uso secondo l'ordine stabilito dalla comunità Share. Tra questi: catalogazione derivata, catalogazione originale, API per ILS, retro-conversione per sistemi di acquisizione e amministrazione locali basati su Formati MARC, reportistica.

Nelle varie fasi di avanzamento le istituzioni partecipanti sono chiamate a collaborare attivamente con la ricchezza dei propri dataset, in modo che le decisioni relative alle fasi successive siano basate su fondamenta solide per un'implementazione su vasta scala nella comunità delle biblioteche, degli archivi e dei musei. I dati utilizzati non vengono semplicemente convertiti in RDF, ma sono anche arricchiti con identificatori e interconnessi, per consentire il loro utilizzo nell'ambiente dei linked data. Questi processi contribuiscono alla costruzione di un modello realistico di attività dell'utente, in modo da evidenziare sia eventuali criticità che potenziali vantaggi del passaggio all'ambiente dei linked data. Share-VDE, sviluppato secondo il principio di funzionalità, fornisce vari ambienti e interfacce per la creazione e l'arricchimento dei dati e offre flussi di lavoro in grado di rispondere alle differenti esigenze dei bibliotecari, professionisti, studiosi, ricercatori e studenti partecipanti.

Casalini Libri (www.casalini.it) è un'agenzia bibliografica che produce record di autorità e bibliografici; fornisce libri e riviste alle biblioteche; offre una varietà di servizi di sviluppo e di sviluppo delle collezioni; lavora in collaborazione con editori e biblioteche nel fornire contenuti elettronici.

@Cult (www.atcult.it) è una società di sviluppo software specializzata nella conversione di dati per LD; fornisce Integrated Library System e strumenti per il discovery, che offrono soluzioni tecnologiche efficaci e innovative per migliorare la gestione delle informazioni e la condivisione delle conoscenze.