

Quando Google non esisteva: i formati bibliografici e SBN

Susanna Peruginelli

Il tema dei formati bibliografici fra ieri ed oggi, e ciò in relazione al Servizio Bibliotecario Nazionale non è un argomento solamente tecnico, ma ha un significato strategico nello sviluppo dei servizi di biblioteca.

E' importante affrontare il tema della strutturazione e codifica dei dati perché sta alla base della cooperazione e condivisione, che è uno dei principi fondanti su cui è stato costruito SBN, ed anche perché l'identificazione precisa dei nostri dati è un requisito fondamentale per metterli in relazione con altri e quindi garantire l'ingresso delle biblioteche nel circuito ormai vastissimo dell'informazione in rete.

La fase progettuale e di iniziale operatività di SBN è stata tutta dedicata alla creazione del catalogo unico e alla sua standardizzazione. Fu stabilito che dovevano essere garantite due funzioni fondamentali del sistema:

- Utilizzo della descrizione dei documenti fatta da altre biblioteche senza duplicare la catalogazione;
- Fornitura tempestiva di informazioni bibliografiche sul posseduto per poi garantire l'accesso ai documenti.

A questo fine è stato necessario stabilire dei requisiti tecnici per la sua operatività e fra questi rientrano precise modalità di registrazione e codifica dei dati che potessero garantire la condivisione dei record bibliografici.

Fu adottato un sistema di basi dati relazionali (una centrale, le altre dei singoli poli) che si basava su un'intuizione con forti connotati di novità per quel tempo, come:

- La scomposizione degli elementi bibliografici che descrivono un documento fin dalla fase di catalogazione;
- Il ricorso ad una struttura di legami fra questi elementi, sempre in fase di catalogazione.

La decisione di non mettere di fronte al catalogatore una maschera di immissione dati in formato MARC ha comportato che il MARC, praticamente l'unico formato di ampio utilizzo allora nelle biblioteche di tutto il mondo, fosse relegato in SBN a ciò per cui è nato, quindi a formato di scambio, facendolo derivare da una transcodifica dei dati, alla loro origine organizzati in una struttura relazionale. Ricordo che Michael Malinconico, noto professore di Library and Information Science presso l'Università dell'Alabama, in visita presso il Gruppo che stava lavorando all'elaborazione e sperimentazione di SBN disse: "ma in questo sistema dove è il MARC?"

Per quanto riguarda i record della Bibliografia nazionale italiana, immessi direttamente in SBN, fu stabilito che fossero poi transcodificati, mediante algoritmi a cui la BNCF ha lavorato intensamente, in UNIMARC per la produzione dei nastri magnetici di scambio al posto del formato ANNAMARC allora in produzione. Grazie a questa decisione è possibile oggi garantire un'uscita di tutti i record SBN, bibliografici e di autorità, in UNIMARC e in MARC21, e quest'ultimo dopo una gestazione piuttosto lunga. UNIMARC in XML è anche il formato utilizzato attualmente nel colloquio fra Indice e poli per la trasmissione dei dati bibliografici.

Richiamo ancora una volta i primi due punti programmatici di SBN, tralasciando gli altri:

1. Ogni elemento bibliografico che costituisce accesso al catalogo è registrato come una unità autonoma (allora non ci è venuto in mente di chiamarla entità!)
2. Le unità autonome comuni a più record bibliografici sono collegate mediante un legame bidirezionale a tutte le unità associate.

Ho insistito su questi due punti per evidenziare come vi furono, nell'impianto di SBN, intuizioni strategiche nuove sull'architettura del catalogo, anticipando idee e funzionalità che hanno poi dato luogo a standard e modelli nati molto dopo, come FRBR-LRM, il modello BIBFRAME messo a punto dalla Biblioteca del Congresso, il linguaggio RDF per la dichiarazione di relazioni fra entità nel web comprensibili senza ambiguità dalle macchine, RDA, tecnologie dei dati connessi, ontologie espresse in termini specifici a singoli settori presenti oggi nel Metadata Open Registry e così via, strumenti con cui si confrontano oggi tutti coloro che, come noi, e sono moltissimi, lavorano per garantire l'accesso all'informazione sul web e alla valorizzazione del proprio patrimonio informativo nella rete, e non sono solo le biblioteche a farlo!

Forse chi ha progettato SBN non immaginava di riuscire concretamente, anche se la speranza era forte, a far lavorare insieme biblioteche così diverse, che hanno contribuito ad un patrimonio ricchissimo, ad oggi più di 14 milioni di titoli con un numero enorme di localizzazioni.

Ma allo stesso tempo non si poteva immaginare la rivoluzione operata dalla rete e lo sviluppo di standard, strutture di dati, linguaggi, software oggi disponibili per identificare in modo preciso i dati e metterli in relazione.

Un requisito importante per il rilancio di SBN oggi è quindi il suo inserimento nel mondo del web, la sua uscita, come si dice oggi con una espressione forse non troppo felice, fuori di sé, per rendere visibile e riutilizzabile in ambienti diversi la ricchezza dei suoi dati e contemporaneamente per utilizzare esso stesso i dati provenienti da fonti diverse.

Gli esempi sono molti a livello internazionale, ed anche in Italia vi sono già alcune iniziative interessanti in questa direzione.

Ricordiamo il sistema avanzato della Bibliothèque nationale di Parigi, Data.bnf, che dà accesso con strumenti di navigazione molto efficaci alle sue collezioni, a quelle digitali di Gallica e in parte a materiale archivistico ricorrendo alle tecnologie dei dati connessi e al linguaggio RDF per la dichiarazione delle relazioni fra entità, tutto questo mediante una conversione da formati preesistenti. E' interessante notare che sono proprio gli esperti della BNF a testimoniare che solo grazie alla granularità del MARC è stato possibile inserire, senza perderne la ricchezza semantica, i propri record nel web e a convertirli in linguaggio RDF presentandoli secondo il modello FRBR.

Segue un breve accenno ad altre importanti iniziative:

- la British National Bibliography offre gran parte dei suoi dati in modalità dati connessi ed anche in formato MARC21;
- OCLC è entrato nel mondo del web con la pubblicazione dei record WorldCat nello schema Schema.org, un vocabolario universale nato dalla collaborazione fra i maggiori motori di ricerca come Google, Yahoo, Bing per arricchire le pagine web correnti di dati strutturati che attribuiscono ad esempio ad un numero o ad una de-

terminata stringa di caratteri un preciso significato. Da notare che all'interno di Schema.org sono state predisposte ontologie per il tipo Creative Works in cui rientrano ad esempio i libri e gli articoli e per il tipo denominato Local Business, anche le biblioteche;

- il portale Bavarikon in Germania,
- il progetto Datos in Spagna;
- BIBFRAME, messo a punto dalla Biblioteca del Congresso in vista di una transizione dal formato MARC ad un modello di dati secondo i principi della famiglia FRBR, FRBRoo, FRAD e FRSAD, che si prefigge l'entrata a pieno titolo delle biblioteche nel mondo del web;
- le tante iniziative del mondo Wikipedia e soprattutto Wikidata, nato nel 2012, una base dati di conoscenza collaborativa, aperta e multilingue, ricchissima di dati strutturati su persone, organizzazioni, argomenti, utilissima per la ricerca e per l'importazione ed esportazione di dati da scambiare con il mondo delle biblioteche.

Anche in Italia non mancano iniziative interessanti, come il progetto SHARE fra università del Mezzogiorno, quello del Museo Galileo che riunisce l'archivio e la biblioteca, quello dell'IBC della Regione Emilia Romagna.

E' importante un accenno al problema della mappatura fra schemi e formati, applicazioni di biblioteca e non, necessaria per far "parlare" mondi diversi e garantire un accesso integrato e la navigazione fra i dati. Molti sono gli strumenti che vengono messi a disposizione, uno degli ultimi, in via di sperimentazione, è MARC21RDF, in cui sono state definite opportune proprietà RDF separate per ogni attributo dei campi fissi MARC e via via per

altri elementi, tanto che nell'Open Metadata Registry, in cui ormai le biblioteche sono largamente presenti, sono ospitati, oltre al vocabolario di ISBD, quello della famiglia FRBR, di RDA, ora la versione sperimentale di MARC21 in RDF, ma anche Biba, Dublin Core Library Application profile, etc..

A questo punto è naturale chiedersi come, di fronte alla proliferazione di schemi di metadati, sia possibile avviare lo sviluppo di sistemi di accesso aperto sul web che garantiscano la presenza delle biblioteche nella rete, sfruttando ed anche collegando esse stesse, mediante l'adozione di identificatori univoci e persistenti, informazioni presenti nel web.

Il significato di mappatura fra formati e schemi di metadati, è cambiato radicalmente: da un approccio basato su database e record siamo passati ad un approccio che deve tener conto di un ambiente aperto, globale, multi dominio, basato sulle tecnologie dei dati connessi. Insomma, dall'approccio basato sul record a quello che si fonda su dichiarazioni, da quello basato sulla sintassi ad uno che ricorre alla semantica, quindi al significato dei dati con cui sono descritti i documenti.

Da qui la necessità, riconosciuta a livello internazionale, di linee guida e buone pratiche che aiutino nelle decisioni sul riutilizzo dei dati, l'estensione di vocabolari e lo sviluppo di nuovi.

Si vuole evidenziare comunque che SBN ha già iniziato "ad uscire fuori di sé. Dal 2009 vengono infatti comunicati in formato UNIMARC i record dell'authority file SBN a VIAF, per un totale, ad oggi, di 200.000 record.

Inoltre, sulla base dell'accordo fra l'Istituto Centrale del Catalogo Unico e Wikimedia Italia, SBN contribuisce i record di autorità a Wikidata, ad oggi con 65.000 voci: questo è

un esempio importante dell'impegno a perseguire una politica progressiva di apertura e pubblicazione dei propri contenuti in rete sotto licenze libere.

In seno all'ICCU è stato anche attivato uno studio per la pubblicazione e l'accesso ai dati SBN secondo la tecnologia dei dati connessi ed un Gruppo di lavoro ha messo a punto la mappatura dei campi di SBN, estratti in UNIMARC dall'OPAC e li ha mappati verso il modello FRBRoo.

Importanti sono le diverse iniziative della Biblioteca Nazionale di Firenze come lo scambio e l'accesso reciproco fra le voci del Nuovo Soggettario e le Subject Headings della Biblioteca del Congresso, la codifica in SKOS dei soggetti del Nuovo Soggettario per garantire una corrispondenza con altri tesauri, la navigazione fra le voci di soggetto e quelle di Wikipedia con accesso all'OPAC, e tutto questo tramite Wikidata.

Oggi è fortemente auspicabile che sia avviata una sperimentazione sistematica di utilizzo di una consistente parte di dati SBN, che offrono tutte le garanzie di qualità per il loro utilizzo in applicazioni di dati connessi, mettendoli a confronto e in relazione con i dati provenienti da fonti diverse, primi fra tutti con quelli di archivi e musei: l'interesse verso applicazioni in questa direzione è senz'altro forte da parte degli utenti e un discreto lavoro di progettazione è già iniziato ormai non solo all'interno dell'ICCU, ma anche da diversi istituti del nostro paese. In particolare in seno al MAB Toscana è stato avviato un lavoro di analisi e raccordo degli standard utilizzati per la descrizione di oggetti e documenti che potrebbe essere utilizzato per l'avvio di sperimentazioni.

Si deve riconoscere comunque come sia essenziale, al di là del necessario raccordo di strumenti e standard fra settori diversi, una

comunità di intenti e volontà di collaborazione a livello istituzionale.

A monte di tutto questo è comunque necessario un consenso su alcuni punti, che sono stati anche il frutto di riflessioni compiute all'interno del Coordinamento MAB Toscana:

- un catalogo, inventario o database non può essere mai completo;
- le relazioni fra risorse informative sono importanti tanto quanto la descrizione di singole risorse;
- metadati aperti di qualità non perfetta hanno un valore importante a fronte di metadati di alta qualità ma chiusi e difficilmente accessibili;
- non esiste un record di metadati ideale che soddisfa tutti gli utenti, ma molti modi di aggregare metadati per rispondere alle mutevoli esigenze di ogni utente.