

TEMI E ANALISI

Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicativi dei dati bibliografici

di Antonella Trombone

*Ogni passo autenticamente cognitivo genera visioni più ampie,
più rigorose e imparziali delle precedenti, ma per effettuarlo
occorrono immaginazione, precisione, e consapevolezza delle mete.*

Alfredo Serrai, 2014

Introduzione

La necessità di analizzare le strutture e le visualizzazioni delle entità bibliografiche nasce dall'esame dei cambiamenti in atto relativi sia ai modelli di codifica dei dati che ai sistemi di recupero dell'informazione. Le nuove tecnologie dei software di catalogazione, basate sul Web, hanno condotto la biblioteconomia a confrontarsi tecnicamente e teoricamente con un modo diverso di condividere dati, documenti e contenuti digitali, e a mettersi in relazione con comunità diverse dei beni culturali, affini ma soggette a regole e standard differenti. Nello stesso tempo, la visualizzazione dei dati delle biblioteche è transitata da una consultazione solo locale a una consultazione basata sul Web, che ha sicuramente enormi capacità comunicative, ma che non segue più solo le regole della biblioteconomia.

La normativa catalografica ha adottato i principi del modello concettuale *Functional requirements of bibliographic records* (FRBR)¹ che sposta l'approccio dall'ogget-

ANTONELLA TROMBONE, Università degli studi della Basilicata, Biblioteca centrale di ateneo; dottoranda presso Sapienza Università di Roma, Dipartimento di scienze documentarie, linguistico-filologiche e geografiche, viale Regina Elena 295, 00161 Roma, e-mail antonella.trombone@uniroma1.it.
Ultima consultazione siti web: 19 gennaio 2015.

1 International Federation of Library Associations and Institutions. Study group on the functional requirements for bibliographic records, *Functional requirements for bibliographic records: final report*. München: Saur, 1998.



to al contenuto della risorsa catalogata andando oltre la materialità dei documenti. Il Web, che evidenzia le relazioni oltre che gli oggetti della catalogazione, sta sperimentando la nuova tecnologia dei linked open data per la pubblicazione dei dati informativi, cambiamento questo destinato a influenzare in maniera non lieve i modelli comunicativi della Library and Information Science².

La pubblicazione dei dati sul Web attraverso la tecnologia dei linked data e il modello di codifica *Resource description framework* (RDF)³ hanno favorito l'avvio di una serie di progetti per la traduzione delle informazioni bibliografiche in dati aperti con lo scopo di renderli parte integrante della struttura del Web e direttamente individuabili dai motori di ricerca. Questi progetti di trasformazione delle registrazioni tradizionali in strutture RDF hanno richiesto un'analisi delle semantiche reali degli elementi della descrizione evidenziando la mancanza di un modello condiviso che ospiti le nuove strutture delle informazioni bibliografiche.

Occorre premettere che questo lavoro si basa, in parte, su studi e sperimentazioni in corso e sulla documentazione prodotta dai gruppi di lavoro ufficialmente coinvolti. I documenti ufficiali e gli studi scientifici pubblicati in merito sono pochi, gli esempi ricavabili dalle sperimentazioni in atto sono ancora poco esaustivi sui modelli che li costituiscono, oppure decodificabili solo in formato macchina.

Il progetto *Bibliographic framework transition initiative*

Presentato dalla Library of Congress il 21 novembre 2012, BIBFRAME, abbreviazione di *Bibliographic framework transition initiative*, è un modello di struttura bibliografica del tipo entità-relazione ancora in stato di bozza e aperto a commenti e revisioni. Il suo scopo è la transizione delle strutture bibliografiche dai formati MARC all'ambiente dei dati collegati e del Web semantico⁴. Il progetto BIBFRAME è preceduto da una serie di studi e rapporti istituzionali, ma anche da appassionati dibattiti riportati dai canali ufficiali della letteratura professionale, tutti relativi al lavoro catalogafico e ai suoi risultati comunicativi.

Il dibattito che precede BIBFRAME

Il documento che è conosciuto come 'rapporto Calhoun'⁵ precede di poco l'enunciazione dei principi sui linked data di Tim Berners-Lee⁶. Pubblicato il 17 marzo 2006,

2 Gli atti del convegno tenutosi nel 2012 presso l'Università di Firenze interamente dedicato alle applicazioni dei linked data nelle biblioteche sono pubblicati in: *Global interoperability and linked data in libraries: special issue*, a cura di Mauro Guerrini, Gianfranco Crupi, Ginevra Peruginelli, «JLIS.it», 4 (2013), n. 1, <<http://leo.cineca.it/index.php/jlis/issue/view/536>>.

3 Per l'applicazione di RDF ai dati delle biblioteche si veda Carlo Bianchini; Mauro Guerrini, *Introduzione a RDA*, Milano: Bibliografica, 2014, paragrafo 3.2.

4 Per un'introduzione al progetto: Mauro Guerrini, *BIBFRAME. Per un nuovo ruolo delle biblioteche nel contesto del Web*, «Digitalia», 9 (2014), n. 1, p. 125-126, <<http://digitalia.sbn.it/article/view/1061/691>>; Mauro Guerrini, *BIBFRAME. Un'ipotesi di ambiente bibliografico nell'era del Web*. In: *Il libro al centro. Percorsi fra le discipline del libro in onore di Marco Santoro. Studi promossi da Rosa Marisa Borraccini, Alberto Petrucciani, Carmela Reale, Paola Zito*, a cura di Carmela Reale. Napoli: Liguori, 2014, p. 103-115.

5 Karen Calhoun, *The changing nature of the catalog and its integration with other discovery tools, final report, March 17, 2006*, Washington, DC: Library of Congress, 2006, <<http://www.loc.gov/catdir/calhoun-report-final.pdf> 38>.

il rapporto commissionato dalla Library of Congress di Washington a Karen Calhoun offre un'analisi sulla situazione esistente e alcune proposte per il miglioramento delle funzioni di ricerca dei cataloghi delle biblioteche. I risultati di un'indagine, attuata anche tramite una serie di interviste, sono schematizzati in un elenco di raccomandazioni, tra cui predomina la richiesta di una maggiore capacità relazionale dei cataloghi, al proprio interno e verso altri strumenti di ricerca. In merito ai dati codificati nei formati MARC il rapporto si rivela abbastanza conservativo, sottolinea l'importanza di preservare la quantità e la qualità dei dati bibliografici codificati tramite i formati MARC e, nello stesso tempo, dichiara la necessità di farli interagire all'interno di un sistema di comunicazione globale. Il dibattito sulle capacità comunicative dei formati MARC in atto in quel periodo è sicuramente uno degli elementi che hanno condotto la Library of Congress a commissionare il rapporto⁷.

La Biblioteca del Congresso nel 2008 pubblica il rapporto *On the record*⁸ che richiama espressamente i punti programmatici del 'rapporto Calhoun'. Nel documento ufficiale si riscontrano continui richiami alla necessità di abbattere i costi della catalogazione; anche lo stesso scopo enunciato dal titolo, il futuro del controllo bibliografico, sembra serva a mettere in relazione gli obiettivi cooperativi delle biblioteche con la necessità di minimizzare i costi delle attività catalografiche⁹. Il rapporto evidenzia come sia il controllo d'autorità che la soggettazione siano due funzioni essenziali, anche se molto costose. In entrambi i casi, l'obiettivo è cercare di contenerne i costi, non l'eliminazione¹⁰. Inoltre, le possibilità di colloquio e di scambio tra diverse basi di dati bibliografiche, non intendendosi strettamente cataloghi online, pone il problema dell'internazionalizzazione dei dati, che richiede l'equivalenza delle intestazioni in lingue diverse. *On the record* afferma che sono necessarie strutture di dati che usino identificatori neutri, non basati sui linguaggi naturali, per i termini e le intestazioni. Il lavoro del Virtual International Authority File viene indicato come esempio che va in questa direzione¹¹. In merito alla soggettazione, il rapporto propone di rendere le Library of Congress subject headings più semplici da applicare e da aggiornare, auspicando la possibilità di aggiungere ai record bibliografici voci di soggetto tratte da altri *thesauri* anche per fornire agli utenti chiavi di accesso più ampie¹².

Il rapporto *On the record* si pronuncia ufficialmente in maniera negativa anche sull'uso dei formati MARC, con diverse motivazioni. In particolare, l'uso esclusivo di tali formati da parte delle comunità bibliotecarie compromette la trasmissione dei dati verso altre comunità dei beni culturali. Comincia a delinearsi la necessità di un formato bibliografico che ospiti i metadati creati dai professionisti, quelli creati dai sistemi di automazione, e, ancora, quelli creati dagli utenti che includono annotazioni, recensioni e

6 Tim Berners-Lee, *Linked data*. 2009, <<http://www.w3.org/DesignIssues/LinkedData.html> 40>.

7 *Library hi tech* nel 2004 dedica due numeri al dibattito sul MARC e sui metadati. Si vedano in particolare Roy Tennant, Karen Coyle in «*Library hi tech*», 22 (2004), n. 1 e 2.

8 Library of Congress working group on the future of bibliographic control, *On the record: report of the Library of Congress working group on the future of bibliographic control*, Washington, DC: Library of Congress, 2008, p. 8, <<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jano8-final.pdf>>.

9 *Ivi*, si veda, ad esempio, p. 1.

10 *Ivi*, p. 19.

11 *Ibidem*.

12 Library of Congress working group on the future of bibliographic control, *On the record* cit., p. 36 e 41.

dati sull'uso delle risorse, e nello stesso tempo li distingue tra di loro. Anche i vocabolari controllati usati dalla comunità bibliotecaria sono identificati da stringhe testuali che servono sia per la visualizzazione che per la gestione informatica dei dati, impedendo lo scambio tra comunità linguistiche diverse e tra diversi sistemi informativi. Il rapporto raccomanda l'uso di sistemi di identificazione non linguistici né testuali¹³.

On the record promuove un piano di realizzazione di FRBR, modello concettuale che, nel periodo di pubblicazione del rapporto, comincia a essere parzialmente applicato in alcuni cataloghi di nuova generazione che sperimentano il raggruppamento delle manifestazioni in base all'opera. La pubblicazione di *On the record* precede quella delle norme catalografiche che adottano il modello concettuale FRBR¹⁴, quindi, a prescindere dalle future standardizzazioni normative, propone uno studio per la sua applicazione al catalogo con l'obiettivo di creare un ambiente bibliografico con elementi e relazioni chiaramente definiti, che possa essere usato in diverse situazioni relative al controllo bibliografico¹⁵.

In maniera implicita il rapporto *On the record* segna il passaggio verso *Bibliographic framework transition initiative* raccomandando un lavoro congiunto della Library of Congress con le comunità interessate per lo sviluppo di una struttura di supporto per i dati bibliografici. La sua pubblicazione provoca reazioni e risposte differenti. La Library of Congress appoggia ufficialmente il rapporto e raccomanda lo sviluppo di un nuovo supporto per i metadati, anche se continua ad affermare che i formati MARC svolgeranno ancora un ruolo importante in campo catalografico¹⁶. Se *On the record* viene criticato su alcuni punti, come gli aspetti relativi alla soggettazione¹⁷, si riconosce anche che esso contiene una serie di raccomandazioni che richiedono cambiamenti sostanziali per le biblioteche, relative alle collezioni, agli strumenti di ricerca, alle evoluzioni cooperative dei flussi di lavoro¹⁸. All'interno della Library of Congress si concretizzano iniziative ufficiali finalizzate alla realizzazione di quanto teorizzato nel rapporto¹⁹.

13 *Ivi*, p. 24.

14 Le *Regole italiane di catalogazione* (2009) e *Resource description and access* (2010).

15 Library of Congress working group on the future of bibliographic control, *On the record* cit., p. 33.

16 Il documento di appoggio ufficiale è: Deanna B. Marcum, *Response to "On the record": report of the Library of Congress working group on the future of bibliographic control*, Washington, DC: Library of Congress, 2008, p. 40, <<http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>> 65>.

17 Thomas Mann, *"The changing nature of the catalog and its integration with other discovery tools, final report, March 17, 2006, prepared for the Library of Congress by Karen Calhoun": a critical review*, Washington, DC: AFSCME 2910, the Library of Congress Professional Guild, 2006, p. 11, <<http://www.guild2910.org/AFSCMECalhounReviewREV.pdf>>. Si ricorda in proposito il commento di Alberto Petrucciani, *La catalogazione, il mercato e la fiera dei luoghi comuni*, «Bollettino AIB», 46 (2006), n. 3, p. 177-185; la traduzione in italiano, sempre di Alberto Petrucciani: Thomas Mann, *Il catalogo e gli altri strumenti di ricerca: un punto di vista dalla Library of Congress*, «Bollettino AIB», 46 (2006), n. 3, p. 186-206.

18 Janet Swan Hill, *Entering an alternate universe: some consequences of implementing recommendations of the Library of Congress working group on the future of bibliographic control*, «Library resources & technical services», 52 (2008), n. 4, p. 218-226.

19 OTR report implementation working group, *"On the record report" Recommendations the Library of Congress should pursue over the next four years: report to the Associate librarian for library services*, Washington, DC: Library of Congress, 2009, <http://www.loc.gov/bibliographic-future/news/OTR_rep_response_final_091509.pdf>.

Nel 2010 OCLC pubblica un documento sul MARC volto a determinare quali elementi del formato siano da considerarsi essenziali per la catalogazione²⁰. L'indagine riporta la frequenza di occorrenza di ogni etichetta MARC in WorldCat e i risultati sono forniti in tabelle dettagliate. Scopo del lavoro è l'individuazione degli elementi essenziali e prioritari del MARC da poter usare per le mappature verso altri schemi di metadati.

Nel 2011 la U.S. RDA test coordinating committee della Library of Congress rilascia un rapporto che raccomanda che *Resource description and access* (RDA), la cui prima edizione si data nel 2010, non sia implementato se non dopo l'aggiornamento della struttura di codifica dei metadati del catalogo, quindi fino a quando non si raggiungano progressi credibili nella sostituzione del MARC²¹. La posizione espressa dalla commissione sembra affermare, quindi, che il solo cambiamento delle regole catalografiche, non collegato ad aggiornamenti allo schema di codifica dei metadati del catalogo, non possa apportare miglioramenti alla sua consultazione²².

Le opinioni sul mantenimento del MARC nel dibattito intercorso prima di BIBFRAME sono discordanti, come dimostrano quella contraria del rapporto *U.S. RDA test* e quella a favore di *On the record* e del 'rapporto Calhoun'. Benché nessuna delle posizioni espresse sembri essere dominante al momento, occorre segnalare che per favorire l'avvio della catalogazione conforme a RDA sono stati definiti i necessari campi del formato bibliografico e di autorità di MARC21. Questo è un segno tangibile che non lascia pensare a un totale abbandono dei formati MARC da parte della comunità angloamericana, almeno non nell'immediato futuro. Per fornire un solo esempio, basta leggere la codifica delle registrazioni catalografiche più recenti della Library of Congress per verificare che gli elementi di *RDA content type, media type* e *carrier type* hanno sostituito le designazioni generale e speciale del materiale e che a quegli elementi sono stati assegnati i nuovi campi 336, 337 e 338 di MARC21²³.

Sempre nel 2011, la Library of Congress comunica ufficialmente le proprie politiche sia in merito alla sperimentazione dei dati collegati che alla creazione di una nuova struttura per i dati bibliografici. Nel primo documento, *Transforming our bibliographic framework*²⁴, che è attualmente disponibile sul sito di BIBFRAME, la necessità di apportare cambiamenti alle strutture bibliografiche e ai modelli di codifica dei dati è motivata solo in parte dalla pubblicazione di *Resource description and access*, in quanto vi ricorrono le stesse motivazioni economiche alla base del rapporto Calhoun e di *On the record* degli anni precedenti. Le istituzioni dei beni culturali devono affrontare tagli a volte

20 Karen Smith-Yoshimura [et al.], *Implications of MARC tag usage on library metadata practices*, Dublin, OH: OCLC Research, 2010, <<https://www.oclc.org/resources/research/publications/library/2010/2010-06.pdf>>.

21 Christopher Cole [et al.], *Report and recommendations of the U.S. RDA test coordinating committee*, Washington, DC: Library of Congress, 2011, <<http://www.loc.gov/bibliographic-future/rda/source/rdatesting-finalreport-20june2011.pdf>>.

22 Si veda in proposito Angela Kroeger, *The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC future*, «Cataloging & classification quarterly», 51 (2013), n. 8, p. 881.

23 Per una storia del MARC aggiornata a RDA si veda Michele Seikel; Thomas Steele, *How MARC has changed: the history of the format and its forthcoming relationship to RDA*, «Technical services quarterly», 28 (2011), n. 3, in particolare p. 328-333.

24 Deanna B. Marcum, *Transforming our bibliographic framework: a statement from the Library of Congress (May 13, 2011)*, Library of Congress, BIBFRAME, General information, <<http://www.loc.gov/bibframe/news/framework-051311.html>>.

drammatici ai bilanci e ciò molto spesso si riflette sulla qualità dei dati che esse producono. La necessità di accrescere il lavoro cooperativo e la condivisione dei metadati è una delle motivazioni che spingono la Biblioteca del Congresso alla sperimentazione di nuove tecnologie, mentre l'occasione formale è fornita dalle necessità applicative di RDA. La sperimentazione della tecnologia dei linked data viene annunciata ufficialmente: essa riguarderà modelli di rappresentazione di dati diversi da quelli MARC ma considerando anche MARC21, in cui sono codificati milioni di record. Tra i modelli da valutare è annoverato anche *Functional Requirements for Bibliographic Records*. In senso innovativo, il progetto si assicura la possibilità di distaccarsi anche solo parzialmente dai modelli funzionali dell'IFLA, perciò la nuova struttura bibliografica potrebbe non basarsi strettamente su FRBR; in senso conservativo, il continuo richiamo alla salvaguardia del patrimonio di dati codificati in MARC21 prova la natura istituzionale del documento, che al bisogno di innovare affianca le cautele per tutelare il patrimonio.

L'avvio del progetto BIBFRAME

L'annuncio del piano generale di BIBFRAME viene dato nell'ottobre 2011²⁵. Il progetto per la nuova struttura bibliografica intende sperimentare i linked data assicurando però la compatibilità con i record MARC esistenti. Tra gli obiettivi di BIBFRAME è evidente l'intento di sostituire MARC21 con una struttura bibliografica che sia un ambiente piuttosto che un formato e che sia agnostica rispetto alle regole catalografiche. La definizione di 'ambiente' bibliografico ricorre anche nel documento ufficiale pubblicato dalla Library of Congress su BIBFRAME, *Bibliographic framework as a web of data*²⁶, e dovrebbe indicare una struttura di dati bibliografici basata sulle relazioni, costruita tecnologicamente sul Web e integrabile anche da contributi esterni alle tradizionali fonti catalografiche. Il documento su BIBFRAME viene firmato da un gruppo di esperti di Zepheira, la società cui la Biblioteca del Congresso ha commissionato lo sviluppo di questa struttura bibliografica, la stessa che si occupa del progetto parallelo Schema.org di OCLC. Il presupposto alla base del progetto è che la comunità delle biblioteche si trovi a dover sviluppare nuovi modelli per i dati bibliografici per la prima volta dopo l'implementazione del formato Machine readable cataloging, la cui ampia diffusione ha segnato un punto di svolta per la catalogazione cooperativa e per lo scambio delle informazioni catalografiche in forma elettronica. Il modello dei dati che si tende a superare ha raggiunto una stabilità e un ruolo di portata storica nell'ambiente delle biblioteche per diffusione e per affidabilità, come dimostrano l'incalcolabile patrimonio internazionale di dati codificati in MARC e la collaborazione catalografica tra le biblioteche che esso ha permesso. La Library of Congress, con le altre istituzioni coinvolte²⁷, propone di lavorare

25 Il documento si trova sul sito di BIBFRAME e contiene anche il piano generale dell'iniziativa della Library of Congress: Deanna B. Marcum, *A bibliographic framework for the digital age*, <<http://www.loc.gov/bibframe/news/framework-103111.html>>.

26 Eric Miller [et al.], *Bibliographic framework as a Web of data: linked data model and supporting services*, Washington, DC: Library of Congress, 2012, p. 3, <<http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf>>; è di recente pubblicazione una traduzione italiana: *Un 'Bibliographic framework' per un Web dei dati. Il 'Linked data model' e i servizi di supporto*, traduzione di Iolanda Cristaldi e Roberto Morelato, «Digitalia», 9 (2014), n. 1, <<http://digitalia.sbn.it/article/view/1066/696>>.

27 Partecipano alla sperimentazione: Library of Congress, National Library of Medicine, Princeton Library, George Washington University Library, Deutsche National Bibliothek, OCLC.

a un nuovo modello su cui fondare il futuro della descrizione bibliografica e da condividere anche con le altre istituzioni dei beni culturali, sebbene, come dimostrato dai vari documenti che precedono l'avvio del progetto, venga sempre ribadito il ruolo istituzionale e conservativo del formato MARC.

Il nuovo modello di struttura bibliografica progettato, o ambiente, si pone l'obiettivo di fornire le basi per il futuro della descrizione bibliografica. I principi su cui esso si fonda sono: la chiara differenziazione dei dati sul contenuto concettuale da quelli sulla manifestazione fisica di un oggetto bibliografico; l'ampio uso di sistemi di controllo d'autorità per identificare in maniera inequivocabile gli elementi informativi; la rappresentazione delle relazioni esistenti tra le entità che ne favorisca l'uso in termini di navigabilità. Allo stesso tempo questo modello, esattamente come il formato MARC, si pone l'obiettivo di consentire ancora lo scambio dei dati tra le biblioteche e deve essere compatibile sia con le nuove strutture di codifica dei metadati che con i nuovi standard di contenuto, includendo esplicitamente *Resource description and access* tra le linee guida cui BIBFRAME fa riferimento²⁸.

Il progetto della Library of Congress trae ispirazione dalla teoria evolutiva del Web di Tim Berners-Lee. Se il potere informativo della Rete è consistito nel fornire ai suoi utenti la possibilità di consultare i documenti, il prossimo passo è focalizzare l'attenzione su cosa essi contengono: i dati. Quel che viene indicato come il 'Web dei dati' comporta un ampliamento teorico e tecnico del concetto di link attuale, estendendo le possibilità ora offerte ai documenti anche a dati tra loro eterogenei, collegando dati codificati informaticamente anche a persone, cose e luoghi. BIBFRAME, sin dal suo stato di bozza, si ispira a questi presupposti teorici studiando un modello di struttura bibliografica sia per le risorse future che per quelle già tradizionalmente codificate in MARC²⁹. Il documento programmatico di BIBFRAME individua il valore dei dati di alta qualità codificati in MARC dalle biblioteche e ipotizza che una loro traduzione nel Web dei dati possa essere utile a diverse comunità collegate ai beni culturali. I concetti di qualità e di valore dei dati codificati in MARC sono sempre associati alla loro caratteristica granularità, ovvero a quella peculiare suddivisione delle informazioni sulle risorse in campi e sotto-campi ricchi, specifici e riusabili in contesti compatibili. Un simile lavoro di modellizzazione comporta una decostruzione e una successiva ricostruzione dell'assetto informativo del MARC, il quale contiene categorie di dati diverse che corrispondono a funzioni diverse cui il formato assolve. Nelle singole registrazioni bibliografiche in MARC coesistono dati sulla creazione intellettuale legata a un'opera, dati sulle sue diverse forme di materializzazione, metadati di tipo informatico o legati alla gestione bibliotecaria delle copie descritte. Analizzando gli elementi delle registrazioni MARC che corrispondono a queste tre funzioni si può attuare una loro riformulazione in termini di informazioni collegabili. La struttura di *Bibliographic framework* è costruita su elementi informativi codificati in MARC e ricollocati in una nuova architettura basata sul Web, dinamica e capace di aggiornare automaticamente le proprie informazioni qualora siano modificate alla fonte. Non è chiaro dal documento se questa affermazione si riferisca solo al patrimonio retrospettivo dei dati codificati in MARC, che dovrebbero confluire in BIBFRAME, oppure se essa implichi che le future procedure catalografiche debbano prevedere che i dati vengano prima immessi, o corretti, in MARC e poi tradotti in BIBFRAME: «Then, as we leverage the Web as an information architecture, whenever

²⁸ *Bibliographic framework as a Web of data cit.*, p. 5.

²⁹ *Bibliographic framework as a Web of data cit.*, p. 6

updates to these MARC Resources are performed (e.g. someone adds new information about a person, new mappings related to a subject, etc.) notification events can occur to automatically update systems that reference these Resources»³⁰. Come nel caso della versione in linked data della Bibliografia nazionale britannica, alla quale si fa riferimento più avanti, da quanto emerge sembra probabile che BIBFRAME costituirà un modello alternativo agli strati catalografici preesistenti.

Le classi e la struttura di BIBFRAME

La struttura di BIBFRAME suddivide i dati in quattro classi. *Creative work* è assimilabile a un elemento catalografico concettuale, un'entità astratta a cui possono essere associate diverse istanze che condividono lo stesso contenuto oppure altre opere, tuttavia individuabile attraverso un punto di controllo uniforme.

Instance rappresenta una materializzazione individuale di *creative work* di BIBFRAME, di natura fisica o digitale, individuabile anch'essa attraverso un identificatore. La relazione tra un'istanza e un'opera in BIBFRAME è univoca.

Authority è la classe delle informazioni d'autorità che sono in relazione con gli elementi delle classi *work* e *instance*.

Annotation è la classe che rappresenta la reale novità di BIBFRAME, in cui confluiscono tre tipologie di dati relativi a *creative work* e a *instance*. Il modello prevede una struttura di annotazione flessibile perché deve contenere dati provenienti da più fonti: professionali, informatiche e fonti non tradizionali. Mentre la gestione del possesso delle biblioteche e il trattamento degli esemplari in fase catalografica³¹ forniscono informazioni sulle copie reali delle manifestazioni accessibili tramite le biblioteche, quindi dati sulla localizzazione e informazioni attinenti particolari esemplari posseduti, la classe *annotation* di BIBFRAME è destinata a ospitare anche informazioni in relazione con la classe *creative work*.

Il primo tipo di annotazioni che sono in relazione con la classe *instance* sono metadati tradizionalmente formulati da catalogatori esperti: dati sulla localizzazione delle risorse, dati sulle riorganizzazioni dei materiali, dati sull'uso e sulle politiche d'accesso alle risorse di un'istituzione. Questi non sono intesi come una caratteristica di un elemento della classe *instance*, ma come un'asserzione, un'annotazione, che una biblioteca fa su quell'elemento.

Il secondo tipo di annotazioni collegabili alla classe *instance* sono quelle prodotte informaticamente, ovvero i metadati creati dai programmi gestionali che identificano e collegano le risorse.

La terza classe di annotazioni entra in relazione prevalentemente con *creative work* e prevede l'inserimento di note create da una terza parte, o importate da fonti attendibili. Le recensioni e i commenti degli utenti dell'istituzione e i dati stessi sul loro uso delle risorse entrano in gioco ufficialmente nella struttura bibliografica insieme ai dati che arricchiscono le informazioni sulle risorse, quali immagini e recensioni tratte da fonti credibili, queste però già entrate nella tradizione catalografica. La tecnologia dei linked data dovrebbe rendere il sistema delle annotazioni di agevole gestione da parte delle istituzioni le quali, sembra evidente, dovrebbero mantenere le funzioni di controllo e di analisi delle informazioni

³⁰ *Ivi*, p. 7.

³¹ In ambito nazionale, per la catalogazione dell'esemplare si vedano le *Regole italiane di catalogazione: REICAT*. Roma: Istituto centrale per il catalogo unico delle biblioteche italiane, 2009, paragrafi 1.2, 14.1.6 e capitoli 7, 20.

inserite dalla 'terza parte'³² della catalogazione. Gli utenti stessi dell'informazione prendono, quindi, parte attivamente al processo catalografico, con un modello che fornisce un riconoscimento teorico e una definizione tecnica del loro possibile apporto.

La classe *annotation* di BIBFRAME è destinata a contenere ed esprimere in maniera chiara anche le informazioni sul ciclo di vita e d'uso di una registrazione bibliografica. Si annoverano tra queste i dati sulle importazioni e sulle condivisioni dei record che determinano le informazioni sulla loro tracciabilità e provenienza. Questo concetto è rapportabile a quello di provenienza di ambito archivistico e museale, anche se BIBFRAME lo definisce un semplice flusso di lavoro che registra l'uso di un record.

La capacità di combinare le informazioni professionali sui prodotti con le note degli utenti è stata fino ad ora una caratteristica commerciale del Web che BIBFRAME cerca di importare in ambito biblioteconomico, confidando sulle capacità delle biblioteche di controllarla e di spostarne a proprio favore l'interesse.

Le iniziative parallele e i rapporti con *Functional requirements for bibliographic records* e *Resource description and access*

Alcuni progetti affini a BIBFRAME, finalizzati alla transizione dai formati MARC a modelli basati sui linked data, hanno contribuito, in varia misura, all'elaborazione del progetto della Library of Congress. La British Library ha sviluppato una versione in linked open data della British National Bibliography³³ che è stata presa a modello da BIBFRAME, in particolare, per la configurazione dei soggetti nella classe *authority*. La British Library ha scelto di non configurare tutti i dati contenuti nei record MARC in RDF/XML; al contrario, ha selezionato per la traduzione in linked data solo alcuni elementi della descrizione, cercando anche di adottare il più possibile vocabolari e termini controllati già esistenti per la nuova versione della Bibliografia nazionale britannica.

La Deutsche Nationalbibliothek³⁴ progetta di mettere a disposizione della comunità del web semantico tutti i propri dati bibliografici, compresi quelli di autorità. I dati configurati in RDF non hanno la stessa complessità dei corrispondenti dati in MARC, mentre sono state definite delle proprietà specifiche basate sul tipo di risorsa bibliografica che BIBFRAME ha adottato. Anche la definizione della classe *authority* di BIBFRAME è stata influenzata dal modello di *authority file* tedesco.

Nell'elaborazione del progetto BIBFRAME la Biblioteca del Congresso riconosce che FRBR e RDA e il loro approccio, definito olistico, al recupero e all'accesso all'informazione sono alla base del modello BIBFRAME³⁵. Né il modello concettuale, né le linee guida forniscono indicazioni sul metodo di codifica o di visualizzazione dei dati; BIBFRAME è, invece, un progetto che sembra finalizzato sia all'elaborazione di un formato che di un modello dei dati, una struttura cioè di visualizzazione degli elementi che potrebbe essere scelta anche dai software per biblioteche. Se RDA non adotta pedissequamente il modello FRBR, come dimostra la sua semplificazione del

32 *Bibliographic framework as a Web of data* cit., p. 11.

33 La maschera di ricerca e tutta la documentazione programmatica sul progetto è accessibile da <<http://bnb.data.bl.uk>>.

34 Si accede ai dettagli del progetto e ai dati disponibili da questa pagina: <<http://www.dnb.de/EN/lds>>.

35 *Bibliographic framework as a Web of data* cit., p. 37.

livello dell'espressione³⁶, rispetto alle linee guida angloamericane BIBFRAME compie un'ulteriore elaborazione del modello concettuale nella definizione delle classi *creative work* e *instance*³⁷. In effetti, su questo punto si può anche sostenere che il progetto della Library of Congress non si limiti a un'interpretazione di FRBR, ma che crei un proprio modello concettuale costituito dalle sue quattro classi – *creative work*, *instance*, *authority*, *annotation* – e dalle loro relazioni che sono rappresentabili tramite grafi piuttosto che da collegamenti gerarchici tra elementi³⁸.

In ogni caso, che si tratti di una semplificazione o di una rielaborazione, l'entità espressione di FRBR in BIBFRAME è logicamente compresa nella classe *creative work*, insieme all'entità opera. Occorre precisare che il documento ufficiale del 2012 che delinea il progetto non è esplicito su questo punto, né lo è il sito ufficiale della Biblioteca del Congresso: in entrambi i casi *creative work* è definita «a resource reflecting a conceptual essence of the cataloging item», mentre *instance* è definita «a resource reflecting an individual, material embodiment of the work»³⁹. Il chiarimento viene fornito sullo stesso sito web della Library of Congress da due documenti di Sally McCallum che definiscono la classe *creative work* di BIBFRAME approssimativamente equivalente a *work* ed *expression* di FRBR, mentre la classe *instance* è destinata a comprendere le manifestazioni di opere ed espressioni⁴⁰. Anche lo studio di OCLC Research sul progetto della Library of Congress interpreta in questo stesso modo la corrispondenza tra le classi di BIBFRAME e le entità di FRBR⁴¹.

Sempre nel 2013, nonostante il chiarimento ufficiale della Biblioteca del Congresso e la traduzione dei campi di MARC21 nelle classi di BIBFRAME comunicata da Sally McCallum, un articolo di Erik Mitchell su *Library technology reports* testimonia come la comunità angloamericana non consideri affatto definita la corrispondenza tra le entità di FRBR, le relazioni di RDA e le classi di BIBFRAME.

36 Si vedano Pat Riva; Chris Oliver, *Evaluation of RDA as an implementation of FRBR and FRAD*, «Cataloging & classification quarterly», 50 (2012), n. 5-7, p. 583; Shoichi Taniguchi, *Viewing RDA from FRBR and FRAD: does RDA represent a different conceptual model?*, ivi, n. 8, p. 934; mi permetto di rinviare a Antonella Trombone, *Da 'catalogare' a 'identificare e collegare'. Riflessioni su Introduzione a RDA*, «Biblioteche oggi», 32 (2014), n. 7, p. 7.

37 Nel caso di BIBFRAME, per il concetto di semplificazione si veda Angela Kroegeer, *The road to BIBFRAME* cit., p. 882; Carlo Bianchini; Mauro Guerrini, *A turning point for catalogs: Ranganathan's possible point of view*, «Cataloging & classification quarterly», 2014, in print, p. 9.

38 Si veda anche Lars G. Svensson, *Are current bibliographic models suitable for the integration with the Web?*, «Information standards quarterly», 25 (2013), n. 4, p. 9.

39 *Bibliographic framework as a Web of data*, Washington, DC: Library of Congress, 2012, p. 8; Library of Congress, *Overview the BIBFRAME model*, <<http://www.loc.gov/bibframe/docs/model.html>>.

40 Sally McCallum, *Bibliographic framework initiative approach for MARC data as linked data* [13 settembre 2012], <<http://www.loc.gov/bibframe/pdf/BFI-IGeLU-ppt2003.pdf>>, slide 23; *Bibliographic framework initiative: data from MARC*, LITA/ALCTS Formats Transition Interest Group [ALA Midwinter, 26 gennaio 2013], presented by Sally McCallum, <http://www.loc.gov/bibframe/pdf/ALAMw2013-lita-alcts-marc-transition_McCallum.pdf>, slide 7 e 10.

41 Carol Jean Godby, *The relationship between BIBFRAME and OCLC's linked data model of bibliographic description: a working paper*, Dublin, OH: OCLC Research, 2013, p. 23, <<http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf>>.

While the Work entity is roughly analogous to the FRBR work concept, retaining the abstract focus of the FRBR concept, the BIBFRAME Instance concept can roughly be described as a conflation between Expression and Manifestation. This has been a focus of discussion within the community as it signals a potential shift away from the delineated FRBR model used in RDA⁴².

Più recentemente Baker, Coyle e Petiya definiscono BIBFRAME come un modello bibliografico costituito da due entità, *work* e *instance*, sostenendo, inoltre, che l'entità *work* di BIBFRAME equivalga approssimativamente alla combinazione delle entità *opera* ed *espressione* di FRBR. Tuttavia, nel testo dello stesso articolo, questi stessi autori ricordano come BIBFRAME sia un progetto tuttora in corso, che potrebbe, quindi, essere sostanzialmente modificato fino al momento della pubblicazione ufficiale dell'ultima versione del modello bibliografico⁴³.

Conclusioni

Cercando di trarre delle conclusioni da quanto esposto, si può sostenere che BIBFRAME potrebbe costituire un modello strutturale per le biblioteche che desiderano tradurre le registrazioni MARC come linked data, anche se occorre tenere presente che gli standard per i dati collegati sono tuttora in rapida evoluzione⁴⁴. Il dibattito sul futuro dei cataloghi delle biblioteche ha evidenziato il problema dell'isolamento dal Web dei loro dati. Le possibili soluzioni indicate sono diverse: l'individuazione di alternative alla codifica dei dati nei formati MARC; la trasformazione dei dati delle biblioteche da descrizioni strutturate e testuali a un insieme di dati singoli, trattabili informaticamente e associabili anche a diversi contesti⁴⁵. Allo stesso tempo, poiché i linked data non sembra possano meramente sostituire i metadati codificati in MARC senza essere incardinati in una struttura teoricamente e applicativamente definita, attorno a questo problema si sta concentrando una consistente attività di ricerca.

Gli standard MARC per la codifica dei dati delle biblioteche sono criticati sia per la loro struttura poco flessibile per un modello relazionale di visualizzazione delle informazioni, sia perché usano stringhe testuali non decodificabili dalle macchine, sia perché non permettono di includere le informazioni aggiuntive che oggi le biblioteche intendono fornire agli utenti. Tuttavia, la stessa letteratura su BIBFRAME riconosce che essi consentono alle biblioteche di conservare un enorme patrimonio di informazioni specifiche la cui ricchezza deve essere salvaguardata. Il progetto della

⁴² Erik T. Mitchell, *Library linked data: research and adoption*, «Library technology reports», 49 (2013), n. 5, p. 27.

⁴³ Thomas Baker; Karen Coyle; Sean Petiya, *Multi-entity models of resource description in the Semantic Web*, «Library hi tech», 32 (2014), n. 4, p. 572, <<http://dx.doi.org/10.1108/LHT-08-2014-0081>>.

⁴⁴ Su dubbi, problemi e opportunità dell'applicazione dei *linked data* ai dati delle biblioteche si segnala Heather Lea Moulaison; Anthony J. Million, *The disruptive qualities of linked data in the library environment: analysis and recommendations*, «Cataloging & classification quarterly», 52 (2014), n. 4, p. 367-387.

⁴⁵ Per i problemi e le soluzioni riportate si vedano Library of Congress working group on the future of bibliographic control, *On the record* cit., p. 24; Mauro Guerrini; Tiziana Possemato, *Linked data: un nuovo alfabeto del Web semantico*, «Biblioteche oggi», 30 (2012), n. 3, p. 11; Karen Coyle, *Changing the nature of library data*, «Library technology reports», 46 (2010), n. 1, p. 14-29; Ead., *Metadata mix and match*, «Information standards quarterly», 21 (2009), n. 1, p. 9-11.

Library of Congress non è ancora chiaro su molti punti, come per esempio la definizione di alcune classi e i flussi di lavoro per le conversioni dal formato MARC21. Soprattutto, proprio esaminando il progetto non si comprende a pieno se effettivamente i formati MARC non saranno più adottati nei nuovi iter catalografici, oppure se BIBFRAME fornirà solo un'alternativa per la ricerca e la navigazione. Nonostante i dubbi esposti, normali se si esamina un processo teorico e i suoi risvolti applicativi ancora in corso di definizione, occorre riconoscere al progetto della Library of Congress il merito di aver promosso la formazione di un'attiva comunità internazionale che lavora non solo sugli schemi di descrizione delle risorse e sui linguaggi di rappresentazione delle entità, ma anche sulle strutture teoriche di rappresentazione dei dati bibliografici, le quali hanno il compito di espletare la funzione comunicativa da sempre propria dei cataloghi. BIBFRAME comincia a considerare i 'dati bibliografici collegati', se si può definirli in questo modo, progettando un contesto per la loro realizzazione comunicativa, un ambiente costituito da nessi logici e sintattici che tengono insieme la struttura dei dati in modo coerente. Usando la definizione di Alberto Salarelli, si tratta di una prima forma di «approccio politico, prima ancora che tecnico»⁴⁶ alla creazione di strutture di linked data di tipo bibliografico.

Articolo proposto il 19 gennaio 2015 e accettato il 18 febbraio 2015.

ABSTRACT

AIB studi, vol. 55 n. 2 (maggio/agosto 2015), p. 215-226. DOI 10.2426/aibstudi-11100

ANTONELLA TROMBONE, Università degli studi della Basilicata, Biblioteca centrale di ateneo; dottoressa presso Sapienza Università di Roma, Dipartimento di scienze documentarie, linguistico-filologiche e geografiche, viale Regina Elena 295, 00161 Roma, e-mail antonella.trombone@uniroma1.it.

Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicativi dei dati bibliografici

Il progetto *Bibliographic framework initiative* (BIBFRAME) della Library of Congress di Washington propone un modello di struttura bibliografica del tipo entità-relazione il cui scopo è la transizione delle strutture bibliografiche dai formati MARC all'ambiente dei dati collegati e del Web semantico. L'articolo analizza il dibattito professionale e scientifico internazionale che conduce a BIBFRAME, le fasi di avvio del progetto, la struttura e le classi del modello proposto, le iniziative e i progetti affini o paralleli a BIBFRAME.

The Library of Congress' *Bibliographic framework initiative* (BIBFRAME)

The *Bibliographic framework initiative* – BIBFRAME – is a project initiated by the Library of Congress in order to design an entities/relationships bibliographic description, capable of easing the transition from MARC to linked data and the semantic Web. The paper describes the professional and scientific debate leading to BIBFRAME, the project's first steps, the structure and classes of the BIBFRAME model, and the other related initiatives.

⁴⁶ Alberto Salarelli, *Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento*, «AIB studi», 54 (2014), n. 2-3, p. 293.