*La biblioteca digitale nelle università italiane:
progetti e prospettive di sviluppo*

Seminario organizzato dai Coordinamenti dei Sistemi bibliotecari delle Università di Genova, Padova e Sassari in collaborazione con la Commissione nazionale Università Ricerca dell'AIB

(Roma, 30 ottobre 2003)

**Il catalogo: uno strumento della tradizione
allo snodo delle reti digitali**

Paul Gabriele Weston
(Università degli studi di Pavia)

Alla diffusione delle risorse elettroniche e delle reti digitali va il merito di aver contribuito a ricollocare il catalogo al centro del dibattito biblioteconomico, dopo almeno due decenni nel corso dei quali l'interesse verso questo strumento della tradizione bibliotecaria era andato progressivamente affievolendosi. Alle elaborazioni teoriche del periodo compreso tra gli incontri di Parigi (1961) e di Copenhagen (1969), sfociate nel varo del programma Controllo Bibliografico Universale, in vista del quale sono stati messi a punto dispositivi catalografici di grande rilievo quali la famiglia degli ISBD e il formato elettronico UNIMARC, sono seguiti quattro o cinque lustri nei quali alle problematiche sui principi e sulle finalità della catalogazione si è sostituito un dibattito, talvolta anche acceso, ma raramente produttivo, sui sistemi informatici e sull'applicazione delle tecnologie alle procedure di catalogazione. In questo modo non hanno trovato terreno fertile le sollecitazioni e le proposte contenute nelle riflessioni di Lubetzky, di Domanovzsky e di Serrai. A partire dalla metà degli anni Novanta, invece, si assiste ad un fiorire di iniziative volte a gettare luce sulle questioni sollevate dalle caratteristiche proprie dei documenti digitali, la loro intangibilità e mutevolezza, il loro trasferimento attraverso reti telematiche sempre più estese e veloci, la costituzione di archivi imponenti e di difficile governo.

In ambito bibliotecario, per rispondere alle esigenze imposte dalla necessità di provvedere alla descrizione della nuova tipologia documentaria, vengono definiti gli ISBD(ER), un insieme di criteri che si collocano, tuttavia, nel filone della tradizione. Più interessante appare, invece, la riformulazione delle definizioni di

pubblicazione monografica e seriale, richiesta dall'affermarsi di nature bibliografiche rese possibili dalle tecnologie più innovative.

Talvolta, per far fronte alle nuove esigenze si ricorre a dispositivi noti già ai tempi del catalogo cartaceo, soluzioni che una ingenua fiducia nella possibilità di risolvere i problemi moltiplicando gli indici e i punti di accesso alle notizie aveva portato a trascurare. E' il caso dei *super records* o intestazioni complesse autore-titolo, teorizzati prima e poi sperimentati da Fattahi, in risposta alla preoccupazione di chi osserva che i cataloghi elettronici, a motivo della propria estensione, presentano evidenti problemi di consultazione nel caso di voci particolarmente ricche di occorrenze, come avviene per gli autori classici, la narrativa e le opere multimediali in genere. L'aggiunta di tali intestazioni complesse alla notizia catalografica permette al sistema di articolare la presentazione dei risultati di una ricerca bibliografica su più schermate successive, all'interno delle quali le notizie vengono raggruppate per titolo normalizzato, per genere, per lingua, per supporto e per anno di pubblicazione [slides]. Si tratta di un dispositivo che, se da un lato risponde all'esigenza di organizzare la consultazione del catalogo in modo tale da non obbligare il lettore allo scorrimento di una lista comprendente alcune centinaia o migliaia di citazioni sommarie, dall'altro lato trova un precedente nelle intestazioni autore-titolo che, specialmente nell'ambito dei cataloghi cartacei di biblioteche specializzate, consente l'immediata individuazione della porzione di schedario relativa alla descrizione degli esemplari di una determinata opera e all'interno di questa definisce l'ordine di inserzione delle schede stesse. La preoccupazione di stabilire criteri efficaci di presentazione dei risultati delle ricerche bibliografiche e di assicurarne un'applicazione il più possibile omogenea fra tutti gli opac in commercio conduce alla redazione di apposite linee-guida che contribuiscono anch'esse ad alimentare la più generale riflessione sulle funzioni del catalogo e sul modo in cui queste possano venire meglio espletate.

Quando al problema della crescita dei cataloghi si incomincia ad aggiungere quello costituito dalla contemporanea presenza di più versioni su supporto differente delle medesime opere e alcune tipologie di biblioteche, in particolare quelle a connotazione specialistica, in seguito alla diffusione dell'editoria elettronica, avvertono l'esigenza di procedere alla costituzione di strumenti

bibliografici ibridi, le questioni in gioco assumono una dimensione tale da richiedere ben più che semplici operazioni di cosmesi al modulo di restituzione dei risultati delle ricerche.

E' a questo punto che prende il via l'elaborazione di quel modello concettuale conosciuto con l'acronimo FRBR. L'universo bibliografico viene analizzato alla luce delle funzioni richieste ad un sistema informativo da parte di tutti coloro che necessitano di dati di tipo catalografico nello svolgimento delle proprie attività. In prospettiva, il numero e la tipologia degli utilizzatori del catalogo aumentano significativamente finendo per comprendere, accanto ovviamente ai bibliotecari e ai lettori, anche i produttori di documenti bibliografici, gli editori, i distributori, i librai, gli enti preposti alla compilazione di repertori e di altri strumenti di ricerca e così via. E' ancora più stimolante l'idea che il catalogo, almeno in alcune sue componenti, possa integrarsi in sistemi di ricerca prodotti da comunità per molti aspetti affini a quella bibliotecaria, ma con la quale sono esistite fino ad oggi barriere difficilmente valicabili, talvolta per pregiudizi o forme di pigrizia mentale, più spesso per l'effettiva difficoltà di far convivere sistemi modellati sulle caratteristiche e sulle esigenze proprie di ciascuna tipologia documentaria. L'identificazione di alcuni concetti e la stesura di un vocabolario sono fra i passi iniziali di quella collaborazione che la realizzazione delle biblioteche digitali e dei sistemi documentari multidisciplinari non può che sottendere. Poiché FRBR, in quanto modello concettuale, non prevede formati di dati o schemi di metadati specificamente definiti, esso ha ispirato altri modelli ideati nell'ambito di comunità differenti da quella bibliotecaria, come <indec>, ABC Harmony e Variations 2 digital music (Indiana University).

Tutte queste opportunità non debbono tuttavia condurci a facili illusioni sulla possibilità di incrementare la qualità complessiva di qualunque catalogo, grazie alla semplice adozione del nuovo modello e al prezzo di qualche procedura automatizzata. In primo luogo occorre ricordare l'indagine statistica condotta da OCLC [*slide*] che ha permesso di stimare attorno al 20 per cento del totale delle notizie presenti in quel catalogo collettivo quelle relative a più manifestazioni della medesima opera. Al tempo stesso è necessario tener conto di alcune aporie segnalate dalla commissione che per conto della comunità bibliotecaria italiana si è preoccupata di effettuare una valutazione del modello, nonché di quelle, niente

affatto trascurabili, ricordate da Serrai in un articolo di recente pubblicazione. Vi è, infine, da valutare la sensatezza di sviluppare un modello nuovo sotto l'impulso delle inadeguatezze del precedente modello e di sostanziarlo poi proprio con quelle notizie conformi al modello che si intendeva rimpiazzare. Se la ragione dei numeri contribuisce a spiegare una simile scelta, il rischio è tuttavia quello di far passare per radicalmente nuovo ciò che invece non è che una riorganizzazione dell'esistente.

In ogni caso, come si diceva, l'applicazione della nuova struttura al catalogo elettronico tradizionale, redatto cioè seguendo i criteri descrittivi prescritti in ISBD (International Standard Bibliographic Description) e con l'applicazione del formato MARC (Machine Readable Cataloguing), è tutt'altro che semplice. Da un lato, infatti, le notizie catalografiche esistenti non comprendono tutte le informazioni richieste. Dall'altro lato, alcuni dati vengono riportati in modo tale da presentare un tasso di ambiguità troppo elevato rispetto a quanto richiesto per l'articolazione dell'informazione bibliografica nei livelli gerarchici previsti dalla nuova struttura, come dimostrano sia la comparazione originariamente effettuata da Tom Delsey, sia la successiva edizione del medesimo studio rielaborata e pubblicata dalla Library of Congress (Functional Analysis of the MARC 21 Bibliographic and Holdings Formats Updated and Revised by the Network Development and MARC Standards Office Library of Congress, September 23, 2003). L'impossibilità di ricorrere a verifiche puntuali sui documenti, dovuta all'estensione raggiunta dai cataloghi delle biblioteche più importanti e dai cataloghi collettivi delle reti, ha limitato le prime sperimentazioni al raggruppamento delle opere, cioè al primo dei livelli gerarchici definiti in FRBR per le entità del primo gruppo. Sono stati, perciò, messi a punto degli algoritmi che, operando all'interno delle componenti tradizionali del catalogo elettronico - l'archivio delle notizie bibliografiche e l'*authority file* - hanno individuato un certo numero di relazioni bibliografiche esistenti tra i documenti rappresentati nel catalogo, organizzando la presentazione dei risultati delle ricerche sulla base di tali relazioni.

Non tutte le suddette sperimentazioni hanno perseguito le medesime finalità. La Library of Congress, ad esempio, ha messo a punto un dispositivo software in grado di effettuare, al momento della singola ricerca e su richiesta dell'utente, il

raggruppamento delle notizie per espressioni e per formato. E' senz'altro la soluzione meno invasiva, in quanto si limita alla realizzazione di una componente, esterna al catalogo, che agisce dinamicamente in fase di restituzione delle notizie bibliografiche, senza richiedere modifiche al loro contenuto o alla struttura complessiva dell'archivio (*FRBR Display Tool*, <<http://www.loc.gov/marc/marc-functional-analysis/tool.html>>). [*slides*]

Nel Gateway for Australian Literature il modello FRBR è stato, invece, implementato in una struttura, assimilabile per certi versi ad un sistema di *authority control*, che si sovrappone al catalogo collettivo nazionale australiano Kinetika. Tale struttura è finalizzata al potenziamento delle funzioni di interrogazione mediante la valorizzazione delle relazioni esistenti tra autori e opere (opere di, a cura di, tradotte da, con prefazione di, recensite da, aventi per soggetto, etc.) e tra autore e autore. Si tratta, quindi, a differenza del progetto della Library of Congress, di una componente effettiva e non virtuale, realizzata cioè mediante l'immissione di informazioni generalmente non presenti nelle notizie bibliografiche. Anche in questo caso, tuttavia, la struttura e il contenuto dell'archivio catalografico sottostante rimangono immutate, in quanto il raggruppamento delle notizie per espressioni e per formato avviene trattando dinamicamente, secondo opportuni algoritmi, in fase di restituzione, i risultati della ricerca lanciata sull'archivio bibliografico principale (*AustLit – Australian Literature Gateway*, <<http://www.austlit.edu.au/>>). [*slide*]

E' più radicale, infine, la scelta operata nel prodotto commerciale VTLS in quanto esso prevede lo scorporo della notizia bibliografica tradizionale in una serie di notizie relative ciascuna ad un livello gerarchico del modello FRBR. Il catalogatore, operando retrospettivamente su blocchi di notizie esistenti o trattando la singola notizia, può produrre registrazioni dotate di una struttura particolare nella quale vengono inserite le informazioni bibliografiche relative allo specifico livello gerarchico (<<http://www.vtls.com/index.shtml>>). [*slides*]

In prospettiva altri sistemi potrebbero implementare il modello FRBR in una delle forme indicate, dal momento che alcune importanti società tra le quali Digibis, Endeavor, Ex Libris, Innovative Interfaces, OCLC e The Library Corporation hanno avviato contatti con il Joint Steering Committee for the Revision of AACR (JSC) Format Variation Working Group nel giugno 2003.

Nel contesto universitario italiano, precipuamente per il versante umanistico, la realizzazione di un sistema bibliografico che implementi il modello FRBR può offrire molteplici opportunità a condizione che si limitino gli interventi retrospettivi mediante procedure automatiche sugli archivi esistenti, rinunciando quindi ad effettuare ricognizioni autoptiche e demandando agli algoritmi di controllo la soluzione alle disomogeneità dei cataloghi, ma vengano privilegiate invece le procedure manuali. Nell'ottica di un catalogo di qualità, realizzato con una accorta distribuzione delle risorse e minimizzando, per quanto possibile, le ridondanze, si può immaginare che i catalogatori ricorrano alle notizie già disponibili in forma elettronica, preferendo ovviamente quelle prodotte da fonti autorevoli ed integrando, ove necessario, quelle di qualità più modesta, e sottopongano quindi le notizie stesse alle procedure occorrenti per dare corpo alla rete di relazioni nella quale si articola il modello FRBR: l'individuazione dei titoli delle opere, la ricostruzione delle famiglie editoriali, l'aggregazione per formati, l'arricchimento bibliografico e così via. Ogni notizia verrebbe in tal modo arricchita con le informazioni necessarie al pieno sviluppo del nuovo modello, senza peraltro che ciò comporti una rinuncia ad applicare con rigore gli standard catalografici adottati in ambito nazionale o internazionale. Va da sé che, dal punto di vista dell'agenzia bibliografica nazionale l'opportunità, davvero unica, di sottoporre una parte cospicua delle notizie presenti nel catalogo collettivo SBN al controllo "documento alla mano" si potrebbe tradurre in un sostanziale contributo alla bonifica del catalogo, con l'obiettivo di ridurre il fenomeno delle notizie duplicate e di intervenire su quelle errate.

L'adesione al modello concettuale espresso in FRBR porta con sé lo sviluppo di altre componenti del sistema bibliografico. Come ricorda Petrucciani, ad esempio, la piena realizzazione della struttura entità-relazioni implica una maggiore focalizzazione sulla descrizione e sulla caratterizzazione delle entità del secondo gruppo, oltre che sull'individuazione del ruolo da esse svolto nella creazione, nella produzione, nella diffusione e nella conservazione dei singoli documenti bibliografici.

Ciò conduce, inevitabilmente, a trattare dei compiti del sistema di *authority control* e dell'importanza che esso affianchi il sistema catalografico con lo scopo primario di fornirgli struttura e coerenza. L'arricchimento del sistema di *authority*

control, ottenuto ampliando la varietà delle informazioni relative alle singole entità, si colloca nell'ottica di cogliere le opportunità offerte dallo strumento elettronico per mettere a disposizione degli utilizzatori nuove funzionalità di ricerca e, al contempo, fa del sistema stesso uno degli elementi essenziali al conseguimento dell'interoperabilità. Dai rapporti con il mondo dell'editoria possono prendere corpo iniziative sul modello di InterParty, un progetto finanziato dalla Commissione Europea che rappresenta un primo esempio di collaborazione nello sviluppo di dispositivi di *authority control* tra rappresentanti dell'industria del libro, istituzioni che operano nel campo della gestione dei diritti e biblioteche, tutte comunità fino ad oggi, per molti versi, estranee l'un l'altra (<<http://www.interparty.org/>>). Trovandosi nella necessità di rappresentare univocamente le entità responsabili per la creazione della proprietà intellettuale, gli editori hanno individuato negli *authority file* una componente fondamentale del sistema di controllo ed hanno perciò coinvolto le istituzioni bibliotecarie in una collaborazione, che potrebbe apportare benefici considerevoli anche a queste ultime.

Il passaggio da un approccio specificamente nazionale ad uno internazionale nel campo della condivisione di risorse è un'altra delle ragioni che spiega il rinnovato interesse verso le problematiche dell'*authority control* e l'esigenza diffusamente avvertita di disporre di un codice internazionale di catalogazione. Le analisi condotte in concomitanza con alcuni progetti europei, quali TEL (The European Library, <<http://www.europeanlibrary.org>>), LEAF (Linking and Exploring Authority Files) e MACS (Multilingual Access to Subjects), hanno mostrato l'incidenza delle differenze nella formulazione dei nomi degli autori personali e degli enti collettivi e nella scelta delle voci di soggetto dovute all'uso delle lingue e dei codici catalografici dei singoli Paesi sulle modalità di effettuazione delle ricerche e sui risultati conseguenti. Favoriti da una serie di fattori, tra i quali l'uso crescente di formati standard di comunicazione e lo sviluppo di sistemi complessi a gestione distribuita, questi progetti mirano a rendere possibile la condivisione di quanto, nell'ambito delle risorse create nei diversi Paesi per consentire il controllo bibliografico e documentario, risulti necessario anche agli altri partner della comunità bibliotecaria. D'altronde proprio in risposta a tale esigenza, la più recente edizione delle linee guida per la creazione di notizie di

authority emanata dall'IFLA (Guidelines for Authority Records and References) introduce il concetto di "intestazione equivalente", prospettando la possibilità che ciascuna registrazione possa comprendere, accanto alle forme varianti, le forme adottate in altre lingue o secondo differenti normative catalografiche per individuare una medesima entità. Le conseguenze dell'applicazione dell'intestazione equivalente, dal punto di vista delle funzioni di un sistema di *authority control*, sono molteplici. Applicato a supporto di un metaopac, ad esempio, un archivio di entità con queste caratteristiche può consentire il superamento delle barriere linguistiche e normative che ostacolerebbero altrimenti la piena integrazione tra le risorse bibliografiche interrogate. Pensando, in prospettiva, alla realizzazione di sistemi bibliografici internazionali che implementino il modello FRBR, l'armonizzazione delle intestazioni è condizione essenziale perché la descrizione delle manifestazioni possa essere eseguita in modo coordinato e una volta per tutte.

Tanto più numerosi diventano i punti di accesso e i criteri di ricerca a disposizione dei lettori, tanto più forte viene avvertita l'esigenza di fornire al lettore il supporto di una adeguata contestualizzazione dei dati. Per migliorare l'efficacia dei sistemi si cerca, da un parte, di accrescere la capacità di recupero delle notizie in esso presenti e, dall'altra parte, di fornire all'utente, in connessione alla rappresentazione del documento, una serie di ulteriori elementi conoscitivi. Entrambe le finalità sono alla base di quelle attività, complementari alla catalogazione, che prendono il nome di *bibliographic enrichment*, ossia arricchimento bibliografico, e che sono state fatte oggetto di attenzione da parte di numerose istituzioni, tra le quali la Library of Congress, ove opera il Bibliographic Enrichment Advisory Team (BEAT <<http://www.loc.gov/catdir/beat/>>).

L'arricchimento della registrazione comporta la creazione di indici elettronici, i cosiddetti *table of contents* (o *toc*), collegati mediante un collegamento ipertestuale alle singole notizie del catalogo o addirittura incorporati in esse. La finalità perseguita è quella di fornire al lettore, ad integrazione della descrizione di un documento, un'adeguata presentazione del contenuto attraverso la visualizzazione del sommario e degli indici. L'iniziativa ha trovato la piena collaborazione da parte degli editori, i quali intravedono nell'accresciuta

ricercabilità delle notizie un evidente sostegno alle loro finalità commerciali. Allo scopo, l'organizzazione EDItEUR, che conta oltre novanta membri appartenenti a diciassette Paesi in tutti i continenti (EDItEUR, <<http://www.editeur.org/>>), ha sviluppato un insieme di metadati particolarmente esteso, il modello ONIX (Online Information eXchange, <<http://www.editeur.org/onix.html>>). Si tratta di un formato ideato con la finalità di distribuire l'informazione elettronica sui prodotti editoriali in primo luogo per favorirne l'acquisizione. Proprio la presenza dei metadati descrittivi facilita l'integrazione delle registrazioni da loro prodotte nei cataloghi delle biblioteche ed aumenta il numero dei percorsi di ricerca, in quanto, una volta che il collegamento sia stato stipulato, il lettore può passare dal catalogo commerciale a quello bibliografico e viceversa senza apparenti discontinuità ed attivare, ove esistano, collegamenti ipertestuali verso recensioni e segnalazioni prodotte da enti di ricerca, da istituzioni universitarie e da altri organismi preposti all'allestimento di prodotti referenziali.

Questo insieme di soluzioni appare il più adeguato a dare visibilità al contenuto di quei materiali, per i quali la descrizione tradizionale si risolve in una informazione assai povera e spesso anche inutile ai fini della ricerca affidata alle parole chiave del titolo proprio, come nel caso dei cataloghi di mostre, degli atti di convegni, delle antologie e delle altre opere a carattere poligrafico. Analoghe finalità hanno condotto allo sviluppo dei metadati e del codice identificativo DOI (<<http://www.doi.org>>) e alla sua integrazione nella struttura catalografica basata sul formato MARC. Ideato per la gestione delle transazioni commerciali, il DOI ha la prerogativa di presentare una stabilità nel tempo che altri identificatori, come gli URL, non hanno e viene, inoltre, indicizzato dai motori di ricerca anche quando la risorsa corrispondente non lo è. Di particolare interesse è la funzione denominata *MultiLinking*, [slides] la quale consente di collegare ad un unico DOI il riferimento automatico ad una serie di servizi che l'editore o il distributore di una risorsa possono decidere di attivare a beneficio degli utilizzatori, come l'accesso immediato al catalogo dell'editore, la consultazione delle recensioni e persino l'acquisto del libro mediante una transazione elettronica.

Per arricchire il contenuto del catalogo, la Library of Congress ricorre anche all'inclusione nelle registrazioni catalografiche di informazioni desunte da fonti bibliografiche autorevoli, in primo luogo quelle prodotte dal proprio personale

specializzato, ampliando considerevolmente il numero dei percorsi a disposizione del lettore, in quanto ai criteri tradizionali dell'opac si aggiungono quelli propri degli strumenti di consultazione e delle guide bibliografiche. La combinazione fra la ricerca per soggetti o descrittori con le indicazioni provenienti dal repertorio e le relazioni attivate tra risorse elettroniche, dati provenienti da scansione digitale e informazione referenziale, fa del catalogo un vero e proprio sistema bibliografico complesso [*slides*].

Una analoga attenzione alle problematiche dell'arricchimento bibliografico si può riscontrare all'interno di consorzi di biblioteche come OCLC (<<http://www.oclc.org/home/>>), o presso fornitori commerciali come Syndetic (<<http://www.syndetics.com/index.htm>>) e Blackwell (<<http://www.blackwell.com/level2/TOC.asp>>).

Neanche le istituzioni italiane sono rimaste inattive. Anzi, l'arricchimento bibliografico è oggetto di una serie di attività della Biblioteca Nazionale Centrale di Firenze, volte a valorizzare le registrazioni contenute nel proprio catalogo e prodotte nell'ambito dalla Bibliografia Nazionale Italiana. In seguito all'attuazione del progetto ArSBni (Arricchimento dei Servizi della Bibliografia Nazionale Italiana, <<http://www.bncf.firenze.sbn.it/progetti/index.html>> è possibile sfogliare in rete un numero progressivamente crescente di pagine, che intendono fornire al lettore le informazioni essenziali per il riconoscimento del documento e la valutazione della congruenza con quanto ricercato. Il lavoro viene condotto su circa 70.000 volumi pubblicati in Italia a partire dal 1994. Sono previsti la riproduzione digitale e il riconoscimento ottico per la ricerca a pieno testo degli indici, dei frontespizi e delle pagine preliminari, secondo lo standard ISO 5963 relativo all'analisi testuale dei documenti. Ciascuna delle immagini associate agli oggetti digitali è accompagnata da un *file* XML contenente i metadati descrittivi e gestionali e precisamente le informazioni bibliografiche, i dati relativi all'entità digitalizzata, i parametri e il contesto di produzione dell'immagine.

Nel dispiegarsi dell'intreccio di riferimenti incrociati tra descrizioni catalografiche e bibliografie, documenti testuali ed audiovisivi, rimandi interni ai documenti e tra documenti diversi, prende finalmente corpo il catalogo ipertestuale.