

Un'esperienza di catalogazione di risorse elettroniche: la Biblioteca italiana telematica

di Alberto Petrucciani, Laura Berni, Stefano Bolelli Gallevi, Simona Turbanti

1. La *Biblioteca italiana telematica (BIT)*

Tra le numerose iniziative di creazione di “biblioteche digitali”, nei vari sensi in cui questa espressione è stata adoperata, l'esperienza della *Biblioteca italiana telematica* realizzata a partire dal 1997 dal CiBit (Centro interuniversitario Biblioteca italiana telematica) ha delle peculiarità di impostazione e ha seguito delle linee di sperimentazione particolari che pensiamo possa essere utile presentare al lettore.

La *BIT* è nata come progetto di ricerca interuniversitario d'interesse nazionale, finalizzato a raccogliere e mettere a disposizione in formato digitale i testi della cultura italiana, senza limiti cronologici o di materia, attraverso il lavoro di una ventina di gruppi di ricerca, in diversi settori disciplinari, di quindici università, consorziate nel CiBit, con sede a Pisa. La ricerca è stata coordinata da Mirko Tavoni (1997-2000) e poi da Amedeo Quondam (2000-2003), con finanziamenti del MURST (poi MIUR), delle singole università, di enti locali e di istituti privati¹.

La *BIT* è una biblioteca digitale di testi (*non* di immagini o, più esattamente, *non* di riproduzioni di documenti testuali, visto che in effetti le immagini nella *BIT* ci sono, quando corredano un testo) della cultura italiana dal Medioevo al Novecento, prevalentemente ma non esclusivamente letterari (sono presenti anche testi filosofici, giuridici, storici, geografici, ecc., oltre a testi di carattere documentario), “di origine controllata”, ossia digitalizzati, codificati e riveduti da *équipes* specializzate e secondo precisi e uniformi criteri filologici, organizzati dentro un'architettura propriamente biblioteconomica. Uno degli assunti di base del progetto, fin dal principio, è stato infatti costitui-

ALBERTO PETRUCCIANI, Dipartimento di storia moderna e contemporanea, piazza Torricelli 3A, 56126 Pisa, e-mail a.petrucciani@stm.unipi.it; LAURA BERNI, laureata in Conservazione dei beni culturali nell'Università di Pisa, e-mail l_berni@hotmail.com; STEFANO BOLELLI GALLEVI, laureando in Scienze archivistiche e biblioteconomiche nell'Università di Pisa e tirocinante presso il Sistema bibliotecario, e-mail s.b.gallevi@libero.it; SIMONA TURBANTI, Sistema bibliotecario, archivistico e museale dell'Università di Pisa, Lungarno Pacinotti 44, 56126 Pisa, e-mail s.turbanti@adm.unipi.it.

¹ Per una presentazione generale del progetto si veda Mirko Tavoni – Alberto Petrucciani, *The Italian Digital Library project (Biblioteca Telematica Italiana)*, in: *The digital library: challenges and solutions for the new millennium: proceedings of an international conference held in Bologna, Italy, June 1999*, Boston Spa: IFLA Offices for UAP & International Lending, 2000, p. 87-93 (anche in italiano a <<http://www.aib.it/aib/commiss/cnur/dlittavon.htm>>); Patrizia Lüperi, *La Biblioteca italiana telematica: in Internet il patrimonio testuale della tradizione culturale italiana*, «Biblioteche oggi», 17 (1999), n. 5, p. 76; Centro interuniversitario

to dalla convinzione che una biblioteca digitale degna di questo nome debba dotarsi di strumenti di gestione e di accesso, e quindi in primo luogo di un catalogo, coerenti con gli standard bibliografici e catalografici accettati a livello nazionale e internazionale.

Lo studio e la sperimentazione degli aspetti biblioteconomici e, più in particolare, della struttura catalografica della *BIT* sono stati affidati a una delle Unità di ricerca dell'Università di Pisa, coordinata da A. Petrucciani², insieme al compito di contribuire all'acquisizione dei testi con una selezione di opere "classiche" della bibliografia, biblioteconomia e bibliologia italiana dal Settecento alla metà del Novecento³.

Le finalità della *BIT*, offerta liberamente in rete a un pubblico di lettori e di interessati ma destinata anche a un impiego di ricerca scientifica, e perciò agli specialisti degli studi storico-linguistici e filologici, richiedevano non solo un attento controllo della qualità dei testi ma anche la disponibilità di un *software* adeguato per la l'analisi e l'interrogazione. Nella prima fase del progetto è stato utilizzato, tramite la stretta collaborazione con l'Istituto di linguistica computazionale del CNR di Pisa, il programma *DBT (Data base testuale)*, sviluppato da Eugenio Picchi e già largamente noto per il suo impiego nella *Letteratura italiana Zanichelli* su CD-ROM e in altre pubblicazioni. Si tratta di un programma sviluppato a stretto contatto con le esigenze della ricerca linguistica e filologica e quindi assai sofisticato – rispetto al periodo in cui è stato elaborato – nelle funzionalità di marcatura, di ricerca, di visualizzazione e di elaborazione. Può essere utile ricordare che le funzionalità necessarie in una banca dati testuale finalizzata anche alla ricerca specialistica vanno oltre la semplice interrogazione per stringhe, con operatori booleani: strumenti importanti sono per esempio la gestione della prossimità (tra più stringhe ricercate) e la produzione di liste di frequenza e concordanze. Per il progetto è stata quindi sviluppata una versione di *DBT Applet Java*, idonea all'impiego in rete senza necessità di installare un programma *client* e con caratteristiche di sicurezza riguardo allo scarico non autorizzato di dati (richieste, per una piccola parte dei testi, dai detentori dei diritti). Parallelamente, per garantire la disponibilità dei testi in formato standard, è stato sviluppato un programma di conversione automatica della marcatura *DBT* nella marcatura *SGML/TEI* (Standard Generalized Markup Language/Text Encoding Initiative), il formato più accreditato per la codifica elettronica di testi in ambiti di ricerca.

Il sito della *BIT* (<<http://cibit.humnet.unipi.it>>) ha operato fino all'autunno 2002 con queste modalità, arrivando a raccogliere e mettere a disposizione circa 1500 testi. Negli anni fra il 1997 e il 2002, però, l'evoluzione delle tecnologie e degli standard, insieme ai limiti e ai ricorrenti problemi tecnici della prima realizzazione, hanno spinto a una riconsiderazione dell'architettura della *BIT*, con l'abbandono del *software DBT* e della relativa codifica e la progettazione di un nuovo sito, con migliori funzionalità e meglio allineato agli standard e agli strumenti di oggi. Per i nuovi testi è stata adottata una codifica *XML/TEI*, avviando nel contempo la conversione del patrimonio testuale pregresso, ed è stata sviluppata una nuova applicazione per la

Biblioteca italiana telematica, *Biblioteca italiana telematica: risultati e prospettive di una ricerca, guida all'uso della biblioteca*, Pisa: CiBit, Centro interuniversitario Biblioteca italiana telematica, 2000. Una "scheda" della *BIT* dal punto di vista dell'utilizzatore si può leggere in Fabio Metitieri - Riccardo Ridi, *Biblioteche in rete: istruzioni per l'uso*, Roma-Bari: Laterza, 2002, p. 171-172.

2 All'attività dell'Unità di ricerca hanno collaborato, oltre agli autori del presente contributo, Simonetta Basso, Matteo Burioni, Susanna Dal Porto, Elisabetta Macchi, Maria Rita Macchi e Vittorio Ponzani. In numerose occasioni ci ha prestato la sua amichevole consulenza Antonio Scolari, che ringraziamo vivamente.

3 Per questo aspetto dell'attività rimandiamo ad Alberto Petrucciani, *Tesoro della biblioteconomia italiana e biblioteca digitale*, «Bollettino AIB», 41 (2001), n. 3, p. 347-350.

ricerca, basata su un motore di concezione innovativa sviluppato nel Dipartimento di informatica dell'Università di Pisa. Il nuovo sito, ancora in fase di allestimento, è liberamente accessibile da maggio 2003 a <<http://dante.di.unipi.it/ricerca/>>. Nel frattempo, per mettere più largamente a disposizione il patrimonio testuale del progetto CiBit e di altre iniziative analoghe minori, è stato attivato dall'aprile 2001 anche un altro sito parallelo e parziale, sostenuto dall'Università di Roma "La Sapienza", con il titolo di *Biblioteca italiana* (<<http://www.bibliotecaitaliana.it>>).

2. Esigenze e problemi del catalogo della BIT

Per la ricerca e la selezione dei testi della BIT, nelle prime fasi del progetto, è stato utilizzato un semplice *database*, gestito localmente con Access e poi riversato in rete, che permettesse l'interrogazione e la creazione di liste per l'utente.

Per comprendere meglio la problematica da affrontare bisogna però chiarire le funzioni, piuttosto sofisticate rispetto a ordinarie raccolte di testi in formato elettronico, che si richiedevano. Semplice, come sempre in questi casi in cui il numero di autori e di titoli trattati è piuttosto contenuto, era la ricerca di un'opera particolare (semmai con qualche problema di controllo nel caso di autori noti con nomi differenti o in forme diverse e di opere con titoli varianti). Ma il catalogo doveva servire anche a una funzione di selezione o di filtro, per individuare insomma un insieme di opere, un *corpus* entro la biblioteca, su cui compiere poi una interrogazione testuale mirata.

Gli elementi rilevanti identificati a questo scopo, data la natura della raccolta e la sua funzione, sono stati essenzialmente di tre tipi:

- la *lingua*, che ovviamente è per lo più, ma non esclusivamente, l'italiano;
- il contesto di origine, attraverso la registrazione del *periodo* di creazione dell'opera (codificato in decenni o, quando non fosse possibile determinarlo più precisamente, in un arco di anni più ampio) e della *località o area geografica* di origine;
- la forma e il genere, attraverso una ripartizione complessiva in tre *categorie* esclusive, Prosa Versi e Prosimetro, e una più dettagliata definizione di *generi letterari e tipologie testuali* diverse.

Questi cinque elementi sono stati ritenuti importanti per poter svolgere un'interrogazione mirata, p.es. sui testi in una particolare lingua o dialetto, di un particolare periodo (o anteriori o posteriori a una certa data), di una certa area geografica, di un certo genere, solo prosastici o solo in versi, naturalmente con la possibilità di usare tutti questi criteri di selezione o esclusione in combinazione fra loro.

Accanto a questi elementi di contestualizzazione e classificazione dei testi, si è rilevata l'utilità di elementi di selezione relativi alle caratteristiche dell'edizione elettronica: la tipologia di edizione del testo (critica, diplomatica, ecc.), la presenza di elaborazioni o marcature particolari (lemmatizzazione, ecc.) e – anche per finalità gestionali – le connesse responsabilità (curatore e Unità di ricerca di provenienza), la data di pubblicazione in rete e altre informazioni minori.

Gli elementi fin qui ricordati richiedono, per una ricerca pienamente efficace, la possibilità di selezioni multiple (p.es. di Prosa e Prosimetro o di due generi letterari affini) e anche, auspicabilmente, di inclusione logica o gerarchica (il singolo decennio entro un arco di tempo più ampio che lo contiene, Pisa entro la Toscana, un genere letterario entro una più ampia tipologia formale, ecc.), oltre alla combinazione di più elementi di tipo diverso. Una piena funzionalità insomma richiede, oltre alla disponibilità di archivi di autorità per gli autori e i titoli, quella di uno o

più thesauri (almeno per i generi letterari e le tipologie testuali).

Queste considerazioni, come vedremo, ci hanno spinto ad orientarci verso l'utilizzazione, per il catalogo della *BIT*, di un software per la gestione catalografica bibliotecaria, che consentisse di superare le limitazioni di un programma di gestione di basi dati *ad hoc*.

Ma accanto alla gestione di un complesso di elementi di selezione particolari, caratteristici delle finalità speciali di una "biblioteca digitale" come questa, la problematica più impegnativa da affrontare per il catalogo della *BIT* – e di maggiore interesse in questa fase del dibattito teorico sulla catalogazione – riguarda la distinzione concettuale fra gli oggetti di tipo diverso che il catalogo stesso dovrebbe rappresentare.

Per i suoi scopi, il catalogo della *BIT* deve assumere come oggetti di descrizione e di possibile ricerca almeno *quattro entità* di tipo piuttosto diverso: in primo luogo l'*Opera*, che si può presumere sia l'oggetto di ricerca prevalente, quindi il *Testo* o i *Testi* che vengono effettivamente messi a disposizione, l'*Edizione di riferimento* utilizzata (in genere un'edizione a stampa, candidata quindi alla descrizione in una registrazione bibliografica "classica"⁴) e l'*Edizione elettronica* realizzata dalla *BIT* stessa (che a sua volta può essere disponibile in più formati – XML, HTML, PDF, ecc. – non del tutto equivalenti). Ciascuna di queste entità, come si può immaginare, è caratterizzata da propri attributi o relazioni, con i quali disegnare percorsi di ricerca diversi, che vanno per esempio dal recupero di opere che risalgono a uno stesso decennio alla selezione delle edizioni elettroniche in cui è stata applicata una stessa tecnica di marcatura dei fenomeni grammaticali.

Questa problematica richiama alla mente, ovviamente, il modello FRBR, che non poteva non costituire un stimolo e un punto di riferimento per la riflessione. Se il concetto di *Opera*, nonostante le difficoltà di definirlo in maniera pienamente convincente e di applicarlo in modo uniforme, non ha comportato nel nostro caso grossi problemi, situazione assai diversa è quella del concetto di *Espressione* in FRBR. Si è cercato di mostrare altrove come la sua definizione in termini di *identità testuale* (identità di parole e lettere e segni di punteggiatura) sia insostenibile sia in teoria che in pratica, oltre che non pertinente agli scopi di un catalogo bibliografico⁵. Questa conclusione emerge anche dalla nostra esperienza, in quanto data la scelta di un certo testo di una certa opera (di un'edizione critica, di un'edizione curata da un dato studioso, di un'edizione originale a stampa, ecc.), è normale che nelle sue diverse materializzazioni (diverse edizioni che si presentino come immutate, ristampe, perfino esemplari differenti, almeno per l'epoca della stampa manuale) si riscontrino differenze testuali, di solito rare e minute ma comunque effettive, e per esempio si introducano o si correggano errori e difetti. In questo senso, scelta una data versione (*Testo*) di un'*Opera*, ciò che la *BIT* garantisce è la fedeltà a una singola e specificata *Edizione di riferimento*, non necessariamente identica alle altre edizioni che si presentano come contenenti lo stesso testo, unitamente – in molti casi – alla indicazione dettagliata degli interventi (di solito emendamenti di refusi e altri errori dell'originale) compiuti nel passaggio dal testo di riferimento all'*Edizione elettronica*. Se anche, insomma, ci limitassimo a considerare il solo testo come sequenza di lettere, segni di punteggiatura e spazi (esclusi aspetti grafici e di impagina-

4 A rigore, dovremmo ulteriormente distinguere fra l'*Edizione di riferimento* e l'*esemplare* o gli *esemplari* effettivamente seguiti o collazionati, non potendosi escludere *a priori*, com'è noto, l'esistenza di varianti fra esemplari di una stessa edizione.

5 Cfr. Isa De Pinedo – Alberto Petrucciani, *Un approccio all'applicazione del modello FRBR alle regole di catalogazione italiane: problemi e possibili soluzioni*, «Bollettino AIB», 42 (2002), n. 3, p. 267-280.

zione di cui pure è molto azzardato asserire che non incidano sul significato del testo), avremmo in questi casi Espressioni diverse, perdendo così ogni utilità organizzativa di questo concetto, se definito in termini di identità.

L'ottica adottata è stata invece concreta e funzionale, in quanto il livello del Testo è stato utilizzato per rappresentare la presenza nella *BIT*, per una certa Opera, di più versioni diverse, ciascuna dotata di propri elementi di individuazione e con una sostanziale corrispondenza testuale (ma non, si è detto, identità di ogni parola e di ogni segno) fra le sue diverse materializzazioni.

Particolare attenzione è stata dedicata all'Edizione di riferimento, in maniera da individuare, con dati bibliografici completi e controllati, la pubblicazione utilizzata per il testo dell'Edizione elettronica. La completezza dei dati e la loro organizzazione secondo gli standard catalografici permette non solo di dare un'informazione bibliografica precisa e affidabile, ma anche di offrire delle possibilità operative interessanti, per esempio quella di lanciare una ricerca automatica di esemplari dell'edizione di riferimento in uno o più cataloghi o di confrontare automaticamente il catalogo delle edizioni di riferimento e quello di una o più biblioteche per creare dei legami che conducano l'utente dalla registrazione bibliografica dell'edizione a stampa al testo della *BIT* liberamente disponibile in rete.

La descrizione dell'Edizione di riferimento ha comportato non poche difficoltà, non solo per l'acquisizione dei dati (poiché le singole unità e i singoli redattori hanno lavorato un po' in tutta Italia e in tempi diversi con risorse proprie, e non esisteva né esiste quindi una raccolta fisica delle pubblicazioni utilizzate per la digitalizzazione), ma soprattutto per l'estrema varietà delle situazioni da trattare. Contrariamente alla comune digitalizzazione (di solito a livello di immagine) di un certo numero di libri o riviste in maniera integrale, nel caso della *BIT* buona parte dei testi proviene da *parti* di pubblicazioni, antiche o moderne: da una singola commedia edita in una raccolta a una fonte storica pubblicata in uno studio erudito in un fascicolo di un periodico (perfino, in alcuni casi, parti di testi non dotate di un proprio titolo: p.es. i brani relativi alla città di Pisa all'interno di diari o relazioni di viaggio). Inoltre, in alcuni casi l'Edizione di riferimento non consiste in una pubblicazione in senso stretto (e, di conseguenza, la parola "edizione", nell'espressione "Edizione di riferimento", va intesa in senso filologico e non in senso bibliografico): si tratta per esempio di testi editi all'interno di tesi di laurea o trascritti direttamente da originali manoscritti.

L'Edizione elettronica della *BIT*, ultimo oggetto da considerare, richiedeva essa pure – nell'ambito della logica di "biblioteca digitale a denominazione d'origine controllata" – una serie di elementi di individuazione e caratterizzazione, a partire dalle responsabilità ad essa relative (il curatore, talvolta il revisore o supervisore, e l'Unità di ricerca nell'ambito della quale si era svolto il lavoro), dalle modalità dell'edizione, sempre in senso filologico (critica, diplomatica, ecc.), alla data di prima pubblicazione (ed eventualmente di revisione, ecc.), oltre ad elementi peculiari delle finalità di ricerca della *BIT* (p.es. il tipo di marcatura grammaticale, semantica, ecc.).

3. Catalogo "a più oggetti" e integrazione fra UNIMARC e XML/TEI

L'impostazione del catalogo adottata nelle prime fasi del progetto, e realizzata nel sito attivo dalla primavera del 1998 al 2002 (con alcune semplificazioni e in maniera non completa, per problemi tecnici e di tempi), è stata riconsiderata, naturalmente, in rapporto all'adozione del formato XML/TEI per i testi.

Il formato prevede infatti che il testo vero e proprio sia accompagnato da un *header* che comprende, oltre ai dati richiesti dal formato stesso, gli elementi informativi normalmente registrati in un catalogo.

La potenza ed elasticità del linguaggio di marcatura non pongono particolari problemi per la registrazione dei dati citati (e di altri non citati di minore rilievo), ma questione diversa è quella di gestire efficacemente in fase di produzione e soprattutto di ricerca questi elementi: insomma avere non semplicemente un insieme di record contenenti questi elementi, ma un vero catalogo con le appropriate funzionalità.

Ci siamo a questo punto resi conto che un moderno e potente sistema di gestione bibliotecaria, che utilizzasse un formato standard (UNIMARC), costituiva una risposta che, da una parte, superava i limiti di un motore di ricerca o di un'applicazione *ad hoc* (tipiche delle "biblioteche digitali") e, dall'altra, risultava particolarmente coerente con un elemento qualificante del progetto complessivo, l'adozione di standard biblioteconomici riconosciuti.

Ci si è orientati quindi, con l'indispensabile e prezioso sostegno del Servizio per il sistema bibliotecario (oggi Sistema bibliotecario, archivistico e museale) dell'Università di Pisa, verso la realizzazione del catalogo della *BIT* con il software Aleph e nel formato UNIMARC, come base dati separata e autonoma dal catalogo collettivo dell'ateneo pisano ma per quanto possibile omogenea ad esso⁶.

Il catalogo, in via di completamento e già consultabile a <<http://sba.adm.unipi.it:4505/ALEPH/RANDOM9334810/start/bit01?>>, permette la ricerca e la selezione dei testi e, quindi, il collegamento al testo completo, tramite la sua URL; per permettere l'interrogazione testuale su un insieme o *corpus* definito dall'utente o risultato della sua ricerca il catalogo verrà collegato, prossimamente, al nuovo motore di ricerca del sito.

Nel disegno del nuovo catalogo due sono state le problematiche principali da affrontare, il rapporto fra registrazione catalogografica e *header* del testo e la rappresentazione delle diverse entità da descrivere (principalmente Edizione di riferimento ed Edizione elettronica).

Poiché i dati erano a nostro avviso più opportunamente prodotti e gestiti in un sistema di catalogazione, piuttosto che con un *editor* XML, e non essendo evidentemente il caso di duplicare l'attività descrittiva (creazione del record UNIMARC e di parte dello *header* TEI), abbiamo compiuto un'analisi comparativa e sviluppato un convertitore finalizzato alla generazione dello *header* dalla registrazione UNIMARC (con l'aggiunta di informazioni fisse o relative a dati di natura diversa, pertinenti alla gestione del formato). L'analisi e la mappatura dei dati per il convertitore UNIMARC/SGML e UNIMARC/XML è stata compiuta da Laura Berni, con la collaborazione per la realizzazione informatica di David De Servi della ClipWare di Firenze. Il programma realizzato, un eseguibile indipendente con interfaccia grafica, consente di effettuare una semplice e rapida conversione sia multipla (da un file di registrazioni UNIMARC a una serie di file XML/TEI corrispondenti a ciascuna registrazione) sia di singole schede, compresi il recupero e l'integrazione in un unico *header* di dati contenuti in più registrazioni UNIMARC legate fra loro (p.es. quelle di un'Edizione elettronica, della relativa Edizione di riferimento non autonoma e della pub-

⁶ L'analisi per l'applicazione di UNIMARC alle risorse della *BIT* è stata svolta principalmente da Simona Turbanti (al cui contributo, che segue, si rimanda per l'indicazione delle collaborazioni e consulenze che hanno reso possibile la sua realizzazione). Un particolare ringraziamento a Renato Tamburini, dirigente del Sistema bibliotecario, archivistico e museale dell'Università di Pisa: senza la sua generosa disponibilità e il suo sostegno il progetto non sarebbe stato avviato né condotto a termine.

blicazione complessiva che la contiene), con possibilità di controllo e di correzione. Oltre all'impiego per i lavori della *BIT*, ci sembra che possa costituire un esempio interessante di come informazioni create secondo standard professionali consolidati in ambito bibliotecario possano essere diffuse e utilizzate in formati diversi, di applicabilità più generale, e per differenti finalità.

La derivazione dello *header* da una registrazione UNIMARC, modo di procedere a nostro avviso più funzionale rispetto alla via opposta di cercare di registrare in un *header* tutti i dati necessari a una efficace gestione catalografica, può avere un aspetto paradossale, in quanto la registrazione catalografica viene prodotta *prima* che sia completato (con il suo *header*) il documento che la registrazione stessa descrive. Anzi, la zona di identificazione del documento e di informazione sulle sue caratteristiche principali, insomma quella analoga al frontespizio e ai preliminari di una pubblicazione a stampa, è generata *dalla* registrazione catalografica, invece di essere rispecchiata *a posteriori* da questa. Insomma, semplificando per evidenziare il paradosso (e mettendo da parte quindi l'osservazione ineccepibile che una registrazione catalografica standard è qualcosa di assai differente da un semplice rispecchiamento dei dati evidenziati da una pubblicazione), il "frontespizio" dell'Edizione elettronica viene generato a partire dalla registrazione catalografica invece che viceversa.

Questo paradosso è possibile, naturalmente, perché si tratta in questo caso di un complesso di documenti elettronici prodotti dalla stessa organizzazione che ne cura il catalogo, non acquisiti dall'esterno e descritti *a posteriori*, com'è il caso ordinario per una biblioteca. Ma si tratta comunque di una situazione oggi di larga applicabilità, perché per un complesso di ragioni tecniche legali e finanziarie oggi molti istituti si trovano nell'esigenza di catalogare risorse elettroniche che essi stessi producono (o di cui comunque controllano la produzione) e, d'altra parte, per risorse elettroniche non proprie, accessibili in rete, ci si attende sempre più spesso che siano i produttori stessi a sviluppare i relativi sistemi di ricerca e di selezione.

Compiuta questa scelta di partenza, come si vedrà più in dettaglio nel contributo che segue, si è posto il problema di come fornire in maniera completa e conveniente all'utente le informazioni non solo sulla risorsa digitale, ma anche sulla sua fonte di riferimento. A questo scopo è stata innanzitutto sperimentata la soluzione dell'"incapsulamento", offerta da UNIMARC, che consente di registrare in un campo appropriato un certo numero di elementi di descrizione di un'altra pubblicazione collegata (per un esempio si rinvia alle pagine seguenti), ma non ne permette una descrizione completa e articolata, quando è opportuno, su più di un livello. Si è perciò optato per un'architettura pienamente sviluppata, in cui sia la pubblicazione a stampa impiegata come Edizione di riferimento sia l'Edizione elettronica CiBit hanno proprie registrazioni complete, legate fra loro, con i dati applicabili che le riguardano. A questa struttura corrisponde un'interfaccia di ricerca disegnata per rispondere alle finalità che si è cercato di esporre in queste pagine.

An experience in cataloguing of electronic resources: the Italian Telematic Library

by Alberto Petrucciani, Laura Berni, Stefano Bolelli Gallevi, Simona Turbanti

The Italian Telematic Library (Biblioteca Italiana Telematica, BIT) came into being in 1997 as an inter-university research project of national interest. Its aim was to make texts of Italian culture available in digital form through the work of about twenty research groups, in different subject fields, of fifteen universities, who had formed the CiBit consortium (Interuniversity Centre for the Italian Telematic Library), with headquarters in Pisa. The BIT is a digital library of texts of Italian culture that date from the Middle Ages to the twentieth century. They are mainly literary texts “of controlled origin”. This means that they are digitized, marked up and controlled by specialized teams following strict philological criteria, organized within a coherent library architecture with bibliographical and cataloguing standards that are accepted at national and international level. The analysis and implementation of library features and, more specifically, of the catalogue structure of the BIT have been entrusted to one of the research units of the University of Pisa, coordinated by A. Petrucciani. The unit was also responsible for the digitization of a selection of “classical” works of Italian bibliography and librarianship from the Eighteenth to the Twentieth Century. Thanks to cooperation with the Institute of Computational Linguistics of the National Research Council (CNR) of Pisa, the DBT (Textual Data Base) programme developed by Eugenio Picchi in the Applet Java version was used in the first stage of the project; for the sake of standardization a software program for the automatic conversion of the DBT coding into SGML/TEI markup was written and tested. The Web site of the Italian Telematic Library with DBT software was accessible until the Autumn of 2002 at the URL <<http://cibit.humnet.unipi.it>>. However, in 2002 the structure of the site was reconsidered in the light of recent developments and it was agreed to change from the DBT software and markup to newer and better options. The new site is accessible since May 2003 at the URL <<http://dante.di.unipi.it/ricerca/>>, with a search engine of new conception developed at the Computer Science Department of the University of Pisa. It was also decided to adopt the XML/TEI markup. Moreover, for the catalogue functions of the BIT (identification and selection of one or more texts with given characteristics, e.g. type, genre, period, etc.) it was realized that there was a need for a modern and powerful library system that would use a standard format (UNIMARC) and that would be able to overcome the limits of a search engine or of an ad hoc database application, typical of most “digital libraries”. The choice of a standard format for bibliographic data is seen as a qualifying element of the BIT programme.

ALBERTO PETRUCCIANI, e-mail a.petrucciani@stm.unipi.it; LAURA BERNI, e-mail l_berni@hotmail.com;
STEFANO BOLELLI GALLEVI, e-mail s.b.gallevi@libero.it; SIMONA TURBANTI, e-mail s.turbanti@adm.unipi.it.