

Dig *Italia*

Numero 0 - **2005**

Rivista del digitale nei beni culturali

ICCU-ROMA

Tra ricerca e catalogo: un nuovo software per la descrizione dei manoscritti in Germania

Gigliola Barbero

Biblioteca Ambrosiana di Milano

«L'informatica non conosce innovazioni concettuali significative da almeno 30 anni» afferma Gianfranco Prini (Università degli Studi di Milano), mentre in una lezione privata discute con ricercatori umanisti sulle applicazioni delle nuove tecnologie ai loro campi di studio. Anche gli strumenti di archiviazione elettronica utilizzabili da storici, storici dell'arte, studiosi di letteratura e bibliologi sostanzialmente non sono cambiati negli ultimi decenni. Le prime forme di trattamento del testo, non dissimili da quelle attuali, risalgono all'inizio degli anni Settanta del secolo scorso; i database più antichi (se ormai dobbiamo parlare di storia dell'informatica ci sia consentito questo aggettivo) vennero poco dopo e ancora oggi sono questi i sistemi fondamentali utilizzati nella registrazione di tutta l'informazione pubblicata e fatta circolare in Rete. Persino i linguaggi di marcatura, che pure hanno caratteristiche molto importanti proprio per gli umanisti, dal punto di vista concettuale non si distinguono significativamente dai database, poiché – spiega ancora Gianfranco Prini – «entrambi i formalismi impongono una struttura predefinita alla rappresentazione dei dati, al fine di consentirne un trattamento efficiente con procedure altrettanto predefinite».¹

Se quindi gli strumenti concettuali sono già definiti, suddivisi tra *word processor* e sistemi di database o di marcatura, nei quali al linguaggio naturale si accompagna un'esplicita strutturazione semantica, ciò che resta da fare nel presente è realizzare applicazioni rispettose dei contenuti che devono esservi rappresentati e usabili nei singoli settori specifici; ossia occorre che inizi (o prosegua là dove è già iniziata) una fase interlocutoria costruttiva tra specialisti delle nuove tecnologie e specialisti dei contenuti. Questa necessità è particolarmente viva nell'ambito della catalogazione dei manoscritti, un settore della conoscenza che vede incontrarsi e scontrarsi competenze e interessi assai diversi, poiché si sviluppa a cavallo tra mondo della ricerca e mondo delle biblioteche. I ricercatori studiano i manoscritti per finalità esclusivamente culturali, per fare avanzare la ricerca in campo storico o letterario e possono dedicare tempo e risorse pressoché infinite a ciascun problema che incontrano; le biblioteche, dall'altra parte, hanno il compito di fornire accessi ai loro fondi mano-

¹ Su database e linguaggi di marcatura utilizzati in ambito umanistico si veda l'importante sintesi di Stefano Vitali, *Passato digitale*, Milano: Bruno Mondadori, 2004.

scritti in tempi accettabili e per questo scopo le competenze tradizionali di paleografi e studiosi dei testi costituiscono soltanto degli strumenti.

I primi di questi attori, gli studiosi, non possono prevedere in anticipo ciò che troveranno al termine delle loro indagini e quindi preferiscono quasi sempre la prosa libera per riassumere e pubblicare i propri risultati. Veronika von Büren (CNRS, Paris), che conserva nel proprio PC privato la più grande biblioteca digitale dedicata ai manoscritti precarolingi, sostiene che non ha più senso costruire database strutturati e che potenti *information retrieval* potrebbero essere sufficienti per interrogare i testi scientifici; a controprova di ciò dall'archivio di dati conservati nella sua macchina emergono informazioni su tutti i manoscritti dei *Codices latini antiquiores* utilizzando un semplice sistema di *search* che permette di interrogare i testi degli articoli e dei cataloghi recuperati dalla studiosa con l'OCR.² Ma soprattutto dimostra l'importanza dell'archiviazione di testi la versione Beta di *Google Print*, in cui appaiono già anche riproduzioni di cataloghi di manoscritti originariamente a stampa.³

I bibliotecari invece, insieme a poche persone specializzate nel lavoro di catalogazione, cercano di creare di volta in volta degli schemi più rigidi (indici, schede) entro cui registrare le informazioni necessarie, in base alle caratteristiche dei fondi da trattare e alle esigenze degli utenti. Essi, che devono attrezzarsi prima per registrare e poi per gestire grandi quantità di dati, non trovano ostacoli all'uso dei database, che anzi facilitano la normalizzazione dei nomi e la loro identificazione, il trattamento dei titoli, degli incipit, delle date e via dicendo.

In questo senso gli studi sui manoscritti pubblicati dai ricercatori e i cataloghi prodotti all'interno delle biblioteche potrebbero essere interpretati rispettivamente come testi e indici di un unico grande catalogo cumulativo, in grado di valorizzare ogni tipo di documentazione esistente.

Il software ManuscriptumXML, prodotto per volontà della Deutsche Forschungsgemeinschaft (DFG) per incrementare la base di dati interrogabile in *Manuscripta mediaevalia*, pretende di rispondere a questa duplice serie di esigenze.⁴ Anche in questo caso nessuna nuova invenzione, ma solo la riscoperta di una caratteristica poco sfruttata dei database, che in verità permette sia di registrare informazioni espresse in prosa, sia di accumulare *thesauri* e dati rigidamente codificati.

ManuscriptumXML è un database inizialmente prodotto con il nome di HiDA3 per i musei e per l'archiviazione di immagini. Nel 1996 il database fu adattato alle esigenze dei manoscritti e fu usato dai catalogatori nelle biblioteche di Berlino, Stuttgart

² È recente la disponibilità di due software gratuiti in grado di ricercare nei file conservati su un PC: Google desktop (<http://desktop.google.it>), che cerca nei testi di email, Word, Excel, Power Point, PDF, e Windows Desktop (<http://desktop.msn.it>).

³ Su Google Print (<http://print.google.it>).

⁴ Sul sito Manuscripta mediaevalia (<http://www.manuscripta-mediaevalia.de>) si vedano Robert Giel, *Manuscripta mediaevalia. Handschriften aus deutschen Bibliotheken im Internet*, «Gazette du livre médiéval», 39 (automne 2001), p. 34-40 e Gigliola Barbero, *Manoscritti & Computer. Manuscripta mediaevalia*, «Biblioteche Oggi», 20/8 (ottobre 2002), p. 100-102.

e Heidelberg, anche se la maggior parte degli studiosi che si dedicano alla catalogazione dei codici medioevali hanno continuato a scrivere i risultati della loro opera di schedatura con *word processor*. Racconta Robert Giel, della Staatsbibliothek di Berlino: «Essendo un database strutturato con un altissimo numero di campi, HiDA3 non fu mai accettato dai catalogatori abituati a dedicare ai manoscritti dei veri e propri testi “epici” elaborati in *Word*. Così alla fine la DFG ha deciso che se il software non era benvenuto e non era diffuso, ciò era da attribuirsi al sistema su cui si basava; allora, in un meeting tra operatori del settore tenutosi a Lipsia del 2003, vennero discusse le linee guida per una nuova applicazione e i risultati di quell’incontro servirono come base per lo sviluppo di ManuscriptumXML, una sorta di HiDA4».

ManuscriptumXML ora è quasi al traguardo e in fase di test. Esso è in grado di leggere tutti i documenti prodotti con HiDA3 e funziona online, direttamente su un server ospitato dal Bildarchiv Foto Marburg. In più permette di importare ed esportare i dati in XML, secondo uno *schema* prodotto appositamente.

Le informazioni registrate nel database sono articolate in diverse entità: il manoscritto, in quanto singolo oggetto, le date, i luoghi geografici, i cataloghi, le notizie iconografiche, le schede di bibliografia, i nomi di persona, i nomi di enti e i soggetti. Ciascuna entità costituisce il contenuto di un tipo di documento e le entità sono organizzate tra loro in maniera gerarchica. Ovviamente l’entità che raccoglie il maggior numero di informazioni è quella corrispondente al manoscritto mentre le altre costituiscono liste di autorità di diversa tipologia.

Ciascun documento prevede diversi tipi di notizie e termina sempre con il nome dell’ente e del professionista che ne sono responsabili. Per esempio ogni data registrata in relazione alla storia di un codice può essere archiviata in forma normalizzata come elemento a sé e può entrare a fare parte di un indice delle date; ManuscriptumXML comprende in questo caso sia un campo per registrare la forma discorsiva della data (per es. “prima metà del XV sec.”) sia un campo per la forma numerica (per es. “1401-1450”). Invece le informazioni relative a un nome di persona comprendono il nome principale nella sua forma accettata, il numero dell’archivio nazionale *Personennamendatei*⁵ (PND), un secondo nome, il genere, lo stato, la professione, la data e il luogo di nascita e di morte, eventuali indicazioni bibliografiche, note libere e di nuovo i nomi degli autori della notizia biografica.

Nel momento in cui si debba registrare una data relativa alla storia di un codice, oppure un nome di persona, è sufficiente richiamarne solo più la forma normalizzata che resterà collegata rispettivamente all’indice delle date e all’indice dei nomi; in più nella scheda del manoscritto alla forma normalizzata della data si accompagna l’informazione relativa al tipo di datazione, così come al nome si affianca la definizione della relazione che intercorre tra il nome e la scheda, per es. nome di autore, nome dell’attuale possessore, nome di un possessore antico.

⁵ Sull’archivio nazionale PND si veda <http://www.ddb.de/standardisierung/normdateien/pnd.htm>.

Ma a parte gli indici, la maggiore novità di ManuscriptumXML, quella che permette a chi lo ha programmato di definirlo un software orientato al testo (*document-oriented* afferma Robert Giel), è nella duplice modalità di inserimento dei dati relativi a ciascuna descrizione di codice.

Infatti oltre a creare le entità manoscritto, nome, data e così via, utilizzando la scheda *Tabellenansicht* suddivisa in campi, molto simile a quella del vecchio HiDA3, nella nuova versione del software il catalogatore può usare un *editor* di testo disponibile in una nuova maschera di inserimento *Druckansicht*. Tale *editor* permette di creare la descrizione di un manoscritto utilizzando dei paragrafi di testo senza limiti significativi di spazio: un paragrafo dedicato alla segnatura, uno destinato alla descrizione esterna, uno dedicato alla storia, uno dedicato alla legatura, uno per i testi. La descrizione che ne risulta è assolutamente identica a quella che si può elaborare in un qualsiasi *word processor*, ma ManuscriptumXML è in grado di mettere in relazione tutte le schede così realizzate agli indici del database, ossia agli indici che vengono costruiti utilizzando la *Tabellenansicht*.

La descrizione in prosa e gli indici infatti sono collegati da un sistema di link semiautomatico: all'interno di un qualsiasi paragrafo in prosa si può selezionare con il mouse una porzione di testo e quindi, con il tasto F12, si può aprire un menù che permette di inserire quella stringa in uno degli indici. Per esempio nel paragrafo dedicato alla descrizione dei contenuti si può selezionare un nome espresso in forma diretta ("Francesco Petrarca") e quindi inserire quella stringa nel campo nomi degli indici indicandovi la forma normalizzata ("Petrarca, Francesco") ed eventuali altre precisazioni. Ovviamente la tabella degli indici può essere anche usata indipendentemente dall'*editor*, per esempio nei casi in cui si volessero indicizzare solo pochi accessi e per progetti finalizzati a obiettivi parziali.

«Testo e indici: può suonare strano – conclude Robert Giel nel parlarci di ManuscriptumXML – che in Germania si stia cercando di realizzare una struttura così vicina a quella dei vecchi volumi a stampa, dal momento che si intende elaborare un nuovo catalogo elettronico. Ma la verità è che siamo giunti alla conclusione che la maggior parte dei catalogatori non si sarebbero mai convinti a lavorare con un database tradizionale, a meno che non adattassimo il sistema alle loro abitudini e alle loro esigenze».

Quindi ancora nessuna nuova scoperta, nessuna nuova invenzione, ma scelte efficaci finalizzate a rispettare i contenuti e chi li produce e soprattutto adatte sia ai progetti di ricerca sia ai lavori di catalogazione. La decisione di accostare agli indici un sistema di scrittura per registrare schede in prosa avvicina molto le funzionalità di ManuscriptumXML a quelle del vero e proprio XML ed evidenzia la somiglianza concettuale tra database e sistemi di marcatura, non rinunciando all'uso di *authority list* che hanno il pregio di rendere omogenei le forme degli accessi all'informazione. Sarà ora responsabilità della Deutsche Forschungsgemeinschaft la diffusione di questo nuovo software e contemporaneamente saranno i catalogatori e gli studiosi a decidere le sorti di questa applicazione e dell'idea positiva dalla quale si è sviluppata.