Il Manuscripts Catalogue della British Library

Uno straordinario archivio di notizie sui fondi manoscritti

Non è un catalogo di prima mano e non descrive minuziosamente tutte le caratteristiche di ciascun manoscritto, secondo gli schemi cui sono abituati molti bibliotecari e ricercatori. Inoltre, del tutto controcorrente rispetto alle prerogative delle nuove tecnologie, non contiene immagini, e anche le pagine da cui si accede a questo strumento di ricerca sono scevre da qualsiasi ornamento.

Eppure l'archivio che la British Library mette a disposizione all'indirizzo <http://molcat.bl.uk>, con una sessantina di cataloghi a stampa recuperati su supporto elettronico e quasi 70.000 record, probabilmente è l'archivio di notizie su fondi manoscritti più ricco al mondo. Le cifre citate, da sole, basterebbero a qualificare il progetto e a misurarne il peso dei contenuti, ma la chiarezza d'intenti dei suoi curatori e la storia dei lavori risultano ancora più eloquenti e istruttivi, soprattutto per chi si trovi ad affrontare, all'interno di altre istituzioni, problemi simili e ricerchi una soluzione realizzabile in tempi accettabili.

Un sunto ampio e preciso del lavoro condotto alla British Library si può leggere nell'articolo di Rachel Stockdale, *head* della catalogazione dei manoscritti, apparso nel "Journal of the Society of Archivists" (21, 2000, 2) con il titolo *The retrospective conversion of the British Library manuscripts' catalogues: a description of the project.*

Tra il 1994 e il 1995 i responsabili del Department of Manuscripts della grande biblioteca inglese decisero di automatizzare i propri cataloghi, che, come è noto, già godevano di un'ampia diffusione su supporto cartaceo. La scelta di avviare questo processo di informatizzazione non fu dettata quindi dalla necessità di una nuova catalogazione o inventariazione. La decisione fu presa invece con l'intento di dare risposta alle mutate esigenze degli utenti, che chiedevano di poter disporre velocemente di tutte le informazioni dei cataloghi attraverso Internet e di poter eseguire le interrogazioni con forme di accesso più sofisticate del semplice indice dei nomi o della sequenza per segnatura, già esistenti e usati nei repertori a stampa. A ciò si aggiunse la volontà dello stesso dipartimento di costituire uno strumento di lavoro passibile di aggiornamento, non statico, in grado di seguire e registrare il progredire della ricerca e delle nuove scoperte.

Lo studio di fattibilità durò circa due anni, a cavallo tra il 1994 e il 1995, al termine dei quali si impose una precisa scelta di metodo: volendo salvare in massa i dati già esistenti, l'unica strada realisticamente percorribile era quella della scansione dei volumi a stampa. Il progetto alternativo, digitare i dati in file di testo, avrebbe richiesto almeno diciassette anni di lavoro e l'impiego di cinque dattilografi – scrive Rachel Stockdale nel suo articolo – mentre il progetto di riconversione automatica ha reso disponibile in Internet l'attuale base di dati in soli tre anni. Ciò ha richiesto l'intervento stabile di due consulenti tecnici e di una squadra di collaboratori oscillante tra i quattro e i dodici elementi, guidati dalla stessa Stockdale.

Cosa è stato fatto in questi tre anni?

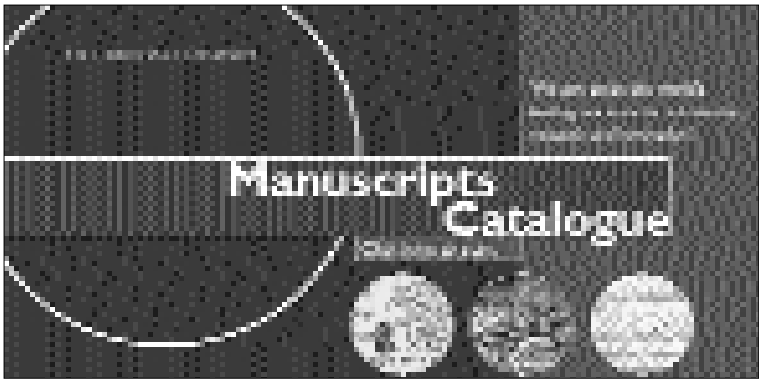
All'inizio ogni pagina di ciascun catalogo è stata "catturata" attraverso uno scanner e memorizzata come un'immagine in formato TIFF. Quindi, usando strumenti software in commercio e senza bisogno di creare applicazioni personalizzate, le immagini sono state trasformate in file di testo attraverso un sistema di Optical Character Recognition (OCR). I file così ottenuti sono stati corretti sia con le funzioni automatiche offerte dallo stesso software di conversione, sia attraverso la lettura eseguita dagli operatori.

Per la seconda fase di lavoro, costituita dal passaggio dal testo libero a un database, è stato invece predisposto un programma specifico. Questo software è stato in grado di suddividere i dati compresi nel testo in base alla loro tipologia, attraverso il rilevamento delle caratteristiche grafiche che nei cataloghi a stampa contraddistinguono le varie porzioni dei cataloghi. Ad esempio, per creare le schede descrittive dei singoli codici, il software è stato programmato affinché leggesse e isolasse il testo compreso tra un numero in grassetto, corrispondente ad una segnatura, e il numero seguente, corrispondente alla segnatura del pezzo seguente.

Con lo stesso metodo sono stati suddivisi anche i dati contenuti nei volumi degli indici che accompagnano i cataloghi descrittivi della British Library. Sempre sulla base delle caratteristiche grafiche e della punteggiatura, ciascun *item* degli indici è stato suddiviso in diversi componenti, tra cui nome, data, lingua.

I dati così articolati dapprima sono serviti a popolare alcuni database (uno per ogni volume) in Access: i dati recuperati dagli indici sono stati allocati in tabelle, mentre le schede recuperate dai volumi descrittivi sono state salvate come oggetti ad esse collegati, grazie alla funzione OLE (Object Linking and Embedding) del database di Microsoft. Quindi i contenuti dei singoli archivi sono stati unificati e trasferiti in SQL Server. Inoltre è stata costruita, appositamente per il Department of Manuscripts, un'interfaccia in Access che permette di intervenire sia sulle descrizioni sia sugli indici, accedendo, ad esempio, attraverso i nomi o le segnature. Tale procedura prevede anche la possibilità di operare sulla lista completa dei nomi, funzione che permetterà nel tempo di elaborare un authority file di forme accettate.

L'interrogazione che gli utenti eseguono oggi attraverso web opera su un archivio che rispecchia fedelmente questa struttura creata dalla riconversione dei dati. Un database SQL Server contiene i record degli indici, mentre, su uno stesso server, le schede descrittive sono salvate come documenti HTML. Per questo motivo l'accesso del pubblico può avveni-



re attraverso due tipi di interrogazione: la prima, qualificata nel frontespizio del catalogo in linea come "Search the Index", opera sul database in cui sono stati archiviati, in formato tabellare, i dati dei volumi a stampa contenenti gli indici; la seconda, quella qualificata come "Search the Descriptions", è in grado di interrogare per parola chiave i file HTML delle schede descrittive. I due tipi di ricerca, operando su archivi diversi, ovviamente non danno sempre gli stessi risultati. Si prenda ad esempio la descrizione del codice Add. 12012:

> Additional Manuscripts 12012
> JUSTINI Epitoma in Trogum Pompeium, sive Historiarum ex Trogo Pompeio, libri XLIV. On vellum, transcribed in the year 1433, by Alessius Germanicus, from a copy belonging to Guarinus Veronensis, and afterwards collated by Martinus Rizonius, disciple of Guarinus. Folio. [12,012.]

L'indice a stampa e il database corrispondente in questo caso rimandano alla scheda di Add. 12012 solo a partire dal nome dell'autore (*Justinus*) e da quello del copista (*Alessius Germanicus*). Gli indici non considerano invece né Guarino Veronese né Martino Rizzoni, implicati nella storia della tradizione del testo. Di conseguenza interrogando il database degli indici con il nome *Guarinus Veronensis*, o con *Martinus Rizonius*, non si ottiene il rimando alla scheda di Add. 12012. Invece questo manoscritto appare tra le risposte se si digita uno dei due nomi tralasciati dagli indici, o parte di essi, nella casella che si apre all'interno della ricerca per parola chiave, quella che agisce sulle schede descrittive archiviate in formato HTML.

Per chi non abbia familiarità con la struttura dei cataloghi a stampa della British Library, questa duplice tipologia di accesso può non risultare intuitiva. Di fronte alle pagine web citate occorre qualche minuto di riflessione e la lettura dell'introduzione è assolutamente necessaria per entrare nella logica dell'archivio e capire come funziona la ricerca. Leggendo la pagina <http:// molcat/bl.uk/ msscat/MSSRETRO.ASP>, infatti, oltre a comprendere che il catalogo costituisce il frutto di una conversione dei precedenti strumenti di ricerca cartacei, si ha notizia esatta di quali tra i volumi pubblicati siano già disponibili on-line, e quali invece manchino ancora all'appello. Ma una volta superato questo lieve pendio introduttivo, la ricerca vera e propria ripaga pienamente della fatica sostenuta. Senza raggiungere biblioteche specializzate, senza lo sforzo di sfogliare contemporaneamente più volumi, senza neanche la fatica di memorizzare di volta in volta i rimandi degli indici, ci si può muovere agilmente tra molti dei fondi della British, verificando ciò di cui si ha bisogno in pochi minuti.

Gli obiettivi del progetto, messi in chiaro al Department of Manuscripts fin dal 1994, sono stati dunque pienamente raggiunti: è stato realizzato uno strumento veloce che permette di condurre diversi tipi di ricerca su un archivio unitario, a partire da qualsiasi computer collegato alla rete.

Costo del progetto: 478.000 sterline. Di queste 358.000 sono state garantite dalla Heritage Lottery Fund, <http://hlf.org.uk>, mentre un sostegno complementare è venuto dalla Gladys Krieble Delmas Foundation, <http://www.delmas.org>.