

Digitalizzazione e catalogazione: quale rapporto?

Separare i due processi per riunirli sotto il segno di una comune interfaccia di dialogo

di Piero Falchetta

Sono ormai innumerevoli, nel mondo degli istituti bibliotecari pubblici e privati, in Italia e all'estero, le esperienze dedicate alla conversione digitale di dati di varia origine e natura. Dalla retroconversione dei cataloghi cartacei, manoscritti o dattiloscritti, alla riproduzione in formato grafico di intere raccolte librerie e non librerie, dalla creazione di imponenti biblioteche digitali per mezzo di processi di trattamento OCR di testi a stampa alla produzione di repertori consultabili tanto in rete quanto su supporti destinati al mercato editoriale, quali i cd-rom, si assiste nel complesso a uno straordinario fiorire di iniziative, progetti e realizzazioni che moltiplicano all'infinito le possibilità di accesso alle fonti del sapere.

Tali progetti e realizzazioni, quanto mai diversificati rispetto ai contenuti, si avvalgono di strumenti tecnologici comuni, in vista innanzitutto del vantaggio che deriva, in termini economici e organizzativi, da processi di trattamento tendenzialmente seriali e standardizzati dei materiali sottoposti alla conversione. Ciò significa, in altre parole, che la stessa piattaforma tecnologi-



L'immagine di un veliero, tratta dal Mappamondo di Fra Mauro (1450 ca), una delle più note opere conservate alla Biblioteca Marciana

ca, costituita dagli stessi strumenti hardware e software e organizzata intorno a un'unica struttura – ancorché aperta – di accesso ai dati e di distribuzione di questi ultimi, può essere impiegata per la realizzazione di archivi digitali paralleli, senza che intervenga la necessità di sostanziali revisioni o adeguamenti dei processi alle diverse tipologie di materiali trattati.

Certo sarà inevitabile porre cure particolari nel trattamento dei materiali antichi e di pregio – mentre altre cure saranno destinate ai materiali librari contemporanei, oggetti di sicuro meno durevoli di quelli realizzati nei secoli scorsi –, sarà

opportuno calibrare la qualità complessiva delle copie digitali così ottenute in relazione alla fungibilità delle stesse nel loro rapporto con l'originale, sarà importante considerare gli aspetti riguardanti la protezione dei dati digitali, soprattutto dal punto di vista del copyright, sarà indispensabile selezionare i documenti da convertire sulla base di scelte determinate da valenze scientifico-documentarie di effettivo spessore culturale, sarà doveroso predisporre canali di accesso che siano, dal punto di vista tecnologico e strutturale, adeguati all'impiego previsto, pianificato o magari anche inaspettatamente occorso una volta messa a regime la base dati. Sono per forza di cose molteplici le variabili delle quali si deve tener conto, per meglio adattarsi, nel corso di tali realizzazioni. In ogni caso, e a partire da qualsivoglia premessa metodologica, l'esito di tale processo sarà omogeneo, in quanto consisterà in un numero n di file contraddistinti da tre caratteristiche peculiari, fra loro determinanti: *durata, riproducibilità, valore tendente allo zero*.

La *durata* dei dati – non quella del supporto nel quale sono archiviati, per la quale occorrerebbero altre considerazioni – è determinata innanzitutto dal loro consistere in sequenze numeriche essenziali, che non sono in via di principio soggette a decadimento nel corso del tempo; tali sequenze di dati sono *riproducibili* all'infinito sui più disparati supporti: listati cartacei, hard disk, cd-rom, DVD, nastri e cartucce, unità di back-up ecc., e ciò indipendentemente o quasi dalla dislocazione fisica dei dati master. Una simile possibilità di moltiplicazione del documento digitale e i bassi costi che tale operazione comporta fanno sì, d'altronde, che il *valore estrinseco* di questi dati sia pressoché nullo, essendo determinato più che altro dagli investimenti tecnologici operati inizialmente per

la conversione, mentre il loro *valore assoluto* consisterà soprattutto nei contenuti dei quali sono portatori.

Dal punto di vista del documento digitale la riproduzione ad altissima risoluzione di un codice di appunti leonardeschi o di qualche celebre libro d'ore miniato e quella, di qualità magari inferiore, di un'annata corrente del "Giornale di Sicilia" non sono sostanzialmente diverse e nessuna di queste è, in termini assoluti, più preziosa o importante dell'altra. Sarà innanzitutto la possibilità di accedere a tali risorse a determinare il loro valore e la loro importanza.

Questa premessa era necessaria per tentare alcune considerazioni sul rapporto fra digitalizzazione e catalogazione all'interno degli istituti bibliotecari, rapporto che sembra a volte presentarsi in termini contraddittori. Le tecniche di digitalizzazione hanno infatti raggiunto uno stato dell'arte abbastanza consolidato che consente, in termini di produttività, una notevole accelerazione dei processi. A semplice titolo d'esempio basterà a tal proposito ricordare come sia stata negli ultimi tempi eseguita, presso la Biblioteca Marciana di Venezia, la digitalizzazione dell'intero fondo delle opere di Vincenzo Coronelli colà conservate, il quale consiste di circa 30.000 pagine manoscritte e a stampa. L'impiego di uno strumento dedicato – nello specifico, uno scanner speciale per libri antichi, capace di riprendere originali fino al formato A1 a una risoluzione massima di 400 punti per pollice – in grado di effettuare 500-600 riprese al giorno, molte delle quali su pagine doppie, ha portato al completamento delle operazioni di conversione digitale di tale fondo entro il termine di un mese circa.

Per la parte grafica del fondo coronelliano, comprendente circa 8.000 incisioni (ritratti, carte geografiche, piante, vedute ecc.), si è invece effettuata una tradizionale ripresa fotografica su positivi 6 x 6 che sono

stati poi passati allo scanner e convertiti in formato digitale. Questa tranche dell'attività, più articolata e complessa, ha richiesto all'incirca sei mesi di lavoro, trascorsi i quali questa notevole quantità di dati ha potuto rendersi disponibile. Tempi paragonabili a questi sono altresì stati necessari per la realizzazione di un altro progetto di digitalizzazione, relativo a circa 7.000 documenti grafici e cartografici, manoscritti e a stampa, di epoca compresa fra il Trecento e il Settecento, mentre tempi ancor più brevi – e ciò grazie all'esperienza e alle tecnologie nel frattempo acquisite – sono infine prevedibili per i nuovi progetti di prossimo avvio.

Il risultato di questa o di altre simili realizzazioni è la costituzione di un archivio digitale composto da decine di migliaia di file, corrispondenti a centinaia di volumi e/o a migliaia di documenti singoli; per accedere a una tale quantità di dati in modo utile, ovvero per poter ricercare e reperire l'informazione voluta, sarà perciò e in ogni caso necessario un accurato lavoro di indicizzazione e di descrizione di tali documenti. Non è tuttavia necessario che l'indicizzazione e la descrizione, per quanto accurate possano queste essere, coincidano con le procedure di catalogazione. La tentazione – lo si è visto, ad esempio, nel caso di molti progetti e attività di digitalizzazione di fondi antichi presentati nel corso del recente convegno padovano dedicato all'argomento (vedi *ultra*) – è infatti spesso, quando si tratti di attività svolte nell'ambito di istituti bibliotecari, quella di approfittare dell'occasione offerta dai progetti di conversione digitale per effettuare la catalogazione standardizzata dei documenti bibliografici convertiti, o per esser più precisi la catalogazione degli originali di cui si è ottenuta copia digitale.

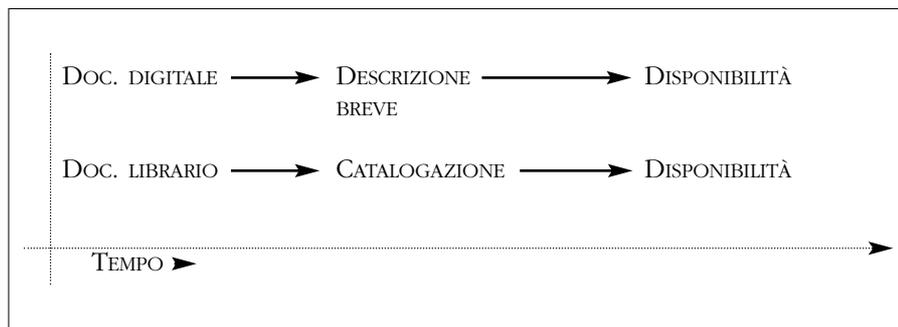
Si verifica a questo punto una discrasia che può avere un effetto pa-

ralizzante sui processi caratteristici del mondo digitale e in particolare su una sua fondamentale proprietà, la pronta disponibilità degli esiti. Se saranno infatti stati necessari alcuni mesi per eseguire la conversione digitale di alcune migliaia di documenti, quanto tempo servirà perché si ottenga la descrizione bibliografica standardizzata degli stessi? E nelle more di tali raggiungimenti, quale potrà essere l'impiego del materiale digitale ottenuto grazie a investimenti – in termini di hardware, software e di servizi interni ed esterni – che si aggirano facilmente attorno alle centinaia di milioni? Non è forse opportuno, se non doveroso, mettere a frutto in tempi quanto più possibile brevi l'impegno economico e organizzativo sostenuto in tali circostanze?

Il rischio di un lungo rinvio della disponibilità dei dati digitali risiede innanzitutto nel forzoso apparentamento fra documento digitale e originale. Non c'è bisogno di ribadire quanto sia laboriosa e complessa la descrizione bibliografica standardizzata dei materiali librari, specie quando si tratti di materiale antico – il più interessato dai progetti di digitalizzazione. La catalogazione è un'attività "scientifica", che richiede ottima preparazione, ampia esperienza, rispetto delle procedure, impegno culturale; essa ha perciò tempi, modi e finalità che coincidono soltanto in piccola parte con quelli caratteristici dei processi di digitalizzazione. Unificare le due attività in modo da far dipendere l'una dall'altra, ovvero considerarle come due momenti complementari di una stessa attività, appare come una contraddizione evidente che ingenera malintesi, ritardi e forte appesantimento dell'attività.

Sarebbe perciò opportuno – nell'assenza di indirizzi precisi o quanto meno di suggerimenti orientativi da parte degli istituti preposti al coordinamento scientifico di tali attività, non ultimo l'ICCU – l'esercizio ➤

Fig. 1 - Schema di riferimento operativo



di una maggior dose di pragmatismo, per mezzo del quale sia possibile restituire ciascuna attività al proprio ambito e alle proprie competenze, ed evitare nel contempo il cattivo impiego delle risorse.

Poiché la digitalizzazione comporta di solito il trattamento di grandi quantità di materiali, sarà comunque necessario, pena l'irreperibilità dei dati, effettuare non soltanto, com'è ovvio, un'accurata indicizzazione degli stessi, ma anche una qualche descrizione che consenta poi normali operazioni di ricerca sui fondi convertiti in digitale. Tale descrizione potrà avere la forma dello *short title*, o potrà comunque limitarsi agli elementi ritenuti indispensabili o ancora a descrivere ciò e quanto le risorse umane impiegate per tale operazioni di descrizione possono consentire in termini di impegno e di capacità. Ciò che le esperienze acquisite fanno infatti apparire con sempre maggior evidenza è la necessità di tenere separati i due processi (conversione digitale e catalogazione), affinché ciascuno di essi possa esprimere al meglio le potenzialità tecniche e "scientifiche" che lo contraddistinguono. Tanto più che tali processi potranno comunque essere riuniti sotto il segno di una comune interfaccia di dialogo, una volta che il più lento e difficile dei due, la catalogazione, abbia esaurito il trattamento bibliografico dei materiali fatti oggetto della conversione digitale.

Alle proprietà indicate per i documenti digitali – *durata, riproducibi-*

lità, valore tendente allo zero – corrispondono infatti le seguenti proprietà dei documenti librari (specie quando si tratti di documenti antichi): *durata*, determinata dalla sperimentata resistenza dei materiali al trascorrere del tempo, unicità, intesa nel senso di esemplare fisico del documento, *alto valore intrinseco ed estrinseco*, riconducibile in senso generale alla nozione storico-filologica di documento. Ed è soltanto al termine delle rispettive procedure, come illustrato dallo schema di riferimento operativo (fig. 1), che dati digitali e dati estrapolati dall'originale potranno essere disponibili simultaneamente, arricchendosi così a vicenda e completando, ciascuno secondo la propria specificità, il portato generale dell'informazione.

Documentazione essenziale

Una vera e propria guida alla creazione di archivi digitali di documenti storici, pubblicata dall'AHDS (Arts and humanities research board, UK) è leggibile all'indirizzo <http://hds.essex.ac.uk/g2gp/digitising_history/index.html> dal quale sono raggiungibili anche importanti risorse digitali quali l'*Oxford text archive*. Altri importanti documenti tecnici sull'acquisizione digitale e la distribuzione di materiali digitali sono i seguenti: *Colorado digitization project*, elaborato dal Colorado Digitization Project Scanning Working Group e consultabile all'indirizzo <<http://coloradodigital.coalliance.org/scanning.html>>, *Digital formats for con-*

tent reproductions, di Carl Fleischhauer, coordinatore tecnico del National digital library program, Library of Congress <<http://memory.loc.gov/ammem/formats.html>>, *NARA Guidelines for digitizing archival materials for electronic access* (National Archives of America) <<http://www.nara.gov/nara/vision/eap/eapspec.html>>, *Creating and distributing high resolution cartographic images*, di David Yehling Allen, Map Collection, Melville Library <<http://www.rlg.org/preserv/diginews/diginews2-4.html#feature>>.

Un interessante documento sugli aspetti tecnici della digitalizzazione dei documenti antichi è stato pubblicato dal Fermi National Accelerator Laboratory: *Digitizing legacy documents: a knowledge base preservation project*. Il documento risente un poco dell'accelerato sviluppo tecnologico, in quanto risale al 1998, ma le indicazioni di fondo qui riportate possono considerarsi senza dubbio ancora valide. Il documento è leggibile al seguente indirizzo: <<http://www.ifla.org/documents/libraries/net/tm-2056.pdf>>.

Una ricchissima bibliografia relativa a tutti gli aspetti dei processi di digitalizzazione (teorici, tecnici, legali ecc.) è consultabile all'indirizzo <<http://info.lib.uh.edu/sepb/toc.htm>>.

Alcuni fra i maggiori progetti di digitalizzazione internazionali sono elencati e descritti a partire dalla pagina <<http://portico.bl.uk/gabriel/bibliotheca-universalis/digit.htm>>. In Italia vi sono molte esperienze in fase di avvio, e altre in stato di realizzazione più o meno avanzato. Fra queste si segnalano i progetti digitali della Biblioteca nazionale centrale di Firenze, all'indirizzo <<http://www.bncf.firenze.sbn.it/progetti/index.html>>, con la presentazione del progetto *Bibliotheca universalis*, del progetto EDEN (Deposito legale di editoria elettronica nazionale), la conversione del catalogo Palatino, comprendente la digitalizzazione delle schede cartacee; la digitalizzazione del fondo galileiano; la retroconversione elettronica del catalogo Sacconi; la presentazione del progetto di restauro virtuale; la digitalizzazione del Catalogo unico delle biblioteche italiane 1886-1957. Il progetto DI.RE. della Biblioteca Braidense di Milano, ancora in fase di avvio, è illustrato all'indirizzo <<http://www.cilea.it/braidens//>

Dire99lv.htm>. I progetti della Biblioteca Marciana di Venezia sono descritti all'indirizzo <<http://marciana.venezia.sbn.it/geo.htm>>; sarà inoltre quanto prima attivato GeoWeb, per mezzo del quale sarà possibile consultare on line la base dati digitale, comprendente circa 16.000 documenti cartografici e stampe di questa biblioteca.

Si segnalano ancora gli atti del recente Convegno "Eredità culturale e nuove tecnologie. La digitalizzazione del libro antico tra accesso e conservazione", Padova, 17 e 18 febbraio 2000, con relazioni sui diversi aspetti ed esperienze di digitalizzazione applicati al libro antico (vedi <<http://www.cab.unipd.it/eventi/antico.php3>>), mentre nel sito

<<http://members.it.tripod.de/diglib/index.html>> sono raccolte numerosissime informazioni e risorse sulle biblioteche digitali. La rivista elettronica "RLG Diginews" (<<http://www.rlg.org/preserv/diginews/>>) contiene infine articoli e interventi dedicati ai diversi aspetti della digitalizzazione applicata ai materiali librari e non librari.