

# SCHEDE UMANISTICHE

Rivista semestrale  
dell'Archivio Umanistico Rinascimentale Bolognese

nuova serie

1996

2

*Estratto*

*Memorandum.*



DIPARTIMENTO DI ITALIANISTICA  
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

  
CLUEB

## *Documentazione e informatica umanistica\**

Dino Buzzetti

Quelle che propongo sono alcune considerazioni sui riflessi che può avere sulla formazione del conservatore documentalista il rapporto fra la documentazione e l'informatica, o meglio il rapporto fra la documentazione e quella che viene solitamente chiamata la 'tecnologia dell'informazione'.

Che il rapporto sia stretto è ovvio e risaputo. «I legami che uniscono comunicazione-documentazione-informazione sono talmente commisti da rendere difficile ogni netta separazione», scrive Paolo Bisogno nella sua non più recente *Teoria della documentazione* (1980), dove si sforza di chiarire la teoria generale, le metodologie e le finalità che fanno della documentazione una «disciplina complessa», non riducibile ad un semplice «ampliamento della biblioteconomia»<sup>1</sup> — un'altra questione su cui occorrerà ritornare. Conviene tuttavia insistere, al di là dell'ovvio, sul rapporto tra documentazione e informatica, e considerare l'intreccio tra lo sviluppo della documentazione e lo sviluppo dei sistemi per l'acquisizione, la conservazione e la trasmissione di elementi di conoscenza.

La prima data importante nella storia della documentazione è certamente quella della fondazione dell'«Office International de Bibliographie», creato circa un secolo fa (1892) da Paul Otlet e Henri Lafontaine. Tre anni più tardi, l'«Institut International de Bibliographie», in cui l'Office si trasforma, intraprende la redazione di un repertorio sistema-

\* Comunicazione presentata al convegno *Università, Beni Culturali, Nuove Professionalità*, organizzato a Ravenna nei giorni 3 e 4 ottobre 1996 per la costituzione della nuova Facoltà di Conservazione dei Beni Culturali.

<sup>1</sup> P. BISOGNO, *Teoria della documentazione*, Milano, Angeli, 1980, pp. 10 e 13.

tico e per autori, realizzato su schede. Nel 1912 fa la sua comparsa il microfilm come mezzo per l'archiviazione dell'informazione in dimensioni ridotte e la Library of Congress è la prima biblioteca al mondo che si dota di un apparecchio di ripresa per i suoi archivi. Nel 1931 comincia ad apparire la parola documentazione e l'Institut International de Bibliographie diventa «Institut International de Documentation». In questi anni vengono scritti i primi lavori teorici sulle tecniche documentali e contemporaneamente si sviluppa l'impiego delle schede intagliate sul margine introdotte dalla McBee Corporation. Nel 1938 dall'Institut International de Documentation nasce la Fédération Internationale de Documentation che si articola in comitati nazionali. In questi anni si hanno i primi studi per l'applicazione alla documentazione delle tecniche delle macchine a schede perforate. A partire dal 1940 queste macchine cominciano ad essere effettivamente impiegate e nello stesso anno si ha la prima realizzazione sperimentale di un selettore fotografico. Il principio delle schede sovrapponibili a selezione visiva era già apparso da tempo (il primo brevetto è del 1915), ma è negli anni immediatamente seguenti la seconda guerra mondiale che l'impiego delle macchine a schede perforate contribuisce notevolmente allo sviluppo dell'automazione documentale. Nel 1950 Calvin Mooers introduce l'espressione *information retrieval* e alla fine degli anni cinquanta, per il trattamento dell'informazione documentale, le macchine a schede perforate sono sostituite dai calcolatori elettronici. Il resto è storia recente. La documentazione è diventata, come si è detto, «una disciplina scientifica con tecniche e metodi proprii», che fanno appello a scienze diverse come la linguistica, la logica, la matematica e l'informatica e si è così giunti alla conclusione che «la documentazione è veramente diventata il trattamento dell'informazione non numerica in tutte le sue forme». <sup>2</sup> Così, la stretta connessione con lo sviluppo delle tecniche per l'archiviazione e la trasmissione di elementi di conoscenza ha progressivamente portato a considerare la documentazione non solo come un'attività concreta, dedicata alla raccolta e alla diffusione dell'informazione bibliografica, ma anche e soprattutto come una disciplina teorica che si occupa in astratto del trattamento dell'informazione.

Allo stesso sviluppo si assiste anche, nel campo della bibliotecono-

<sup>2</sup> J. CHAUMIER, *Les techniques documentaires*, Paris, Presses Universitaires de France, 1971, pp. 9-10.

mia, per quanto attiene all'attività di catalogazione. Considerando sempre più astrattamente la rappresentazione catalografica come trattamento analitico dell'informazione contenuta nel documento, Alfredo Serrai giunge a scrivere:

Documentazione e biblioteconomia [...] si trovano ad affrontare gli stessi fenomeni e gli stessi problemi [...]. In questo senso, è appropriato sostenere che la biblioteconomia, in quanto dottrina dell'organizzazione di documenti rintracciabili per quello che contengono, e perciò come dottrina dei requisiti necessari alla organizzazione delle memorie documentarie, al loro accesso e alla loro utilizzazione — e nella accezione più larga quindi come dottrina della consultazione e della trasmissione delle informazioni registrate — possa per legittimità storica comprendere la documentazione e lo *information retrieval*.<sup>3</sup>

Ora, al di là delle discutibili affermazioni sui reciproci rapporti tra le discipline, qui importa sottolineare che lo sviluppo delle tecniche di amministrazione e di accesso all'informazione registrata ha portato, sia nel campo della documentazione, sia nel campo della biblioteconomia, ad una considerazione sempre più astratta dei problemi riguardanti il trattamento dell'informazione e quindi ad una progressiva convergenza delle norme e delle procedure analitiche adottate ufficialmente, nei due campi, in sede nazionale ed internazionale. In un manuale recente, che reca il titolo *Documentazione e biblioteconomia* e che intende significativamente rivolgersi sia ai «servizi di informazione» sia «alle biblioteche speciali italiane», si segnala ripetutamente che l'attività delle biblioteche e degli istituti per le bibliografie nazionali da un lato, e l'attività dei servizi secondari di indicizzazione e di analisi dall'altro, attraversano «una delicata fase di sovrapposizione». <sup>4</sup>

Se ora consideriamo, in questa prospettiva, i più recenti sviluppi della tecnologia dell'informazione, e soprattutto i progressi della telematica, ci possiamo rendere conto che essi hanno contribuito a rendere ancora più labili i confini tra documentazione e catalogazione. Da una parte, l'accesso a distanza all'informazione catalografica ha attenuato il peso della sua dipendenza dalle attività biblioteconomiche di inventaria-

<sup>3</sup> A. SERRAI, *Guida alla biblioteconomia*, Firenze, Sansoni, 1981, p. 116.

<sup>4</sup> *Documentazione e biblioteconomia. Manuale per i servizi di informazione e le biblioteche speciali italiane*, a cura di M. P. Carosella e M. Valenti, Milano, Angeli, 1989, p. 100.

zione, collocazione e accesso materiale ai documenti, spostando l'accento più sulla rappresentazione analitica del loro contenuto; dall'altra, l'attività di documentazione e di informazione bibliografica ha potuto collegarsi in modo più diretto ed efficace con l'attività di ricerca e di localizzazione del documento. Ma un ultimo, fondamentale passo è stato compiuto con la possibilità di accedere, localmente o a distanza, alla rappresentazione digitale dell'intero documento, sia sotto forma di immagine, sia sotto forma di trascrizione del testo. Mi riferisco alla creazione di grandi archivi digitali di immagini e di testi, ai quali è ora possibile accedere anche a distanza attraverso quel sistema informativo globale, distribuito sull'intero sistema delle reti collegate in Internet, che si è formato negli ultimi anni grazie al vertiginoso sviluppo del cosiddetto WWW o World Wide Web. Ora, proprio questo sviluppo pare imporre un'ulteriore e più specifica riqualificazione teorica della documentazione, con riflessi notevoli sui contenuti fondamentali della formazione.

Ma vediamo attraverso alcuni esempi significativi, che riguardano direttamente la gestione degli archivi e delle biblioteche, qual è la natura specifica dei problemi che si pongono. Citerò solo due casi illustri, che fanno capo alla nuova Bibliothèque de France, la TGB o Très Grande Bibliothèque voluta da Mitterrand, e all'Archivo General de Indias di Siviglia. Alla nuova Bibliothèque de France è prevista l'installazione di postazioni di lettura, accessibili anche a distanza, per la consultazione di vasti fondi di materiali digitalizzati, comprendenti non solo testi e immagini fisse, ma anche suoni e immagini in movimento.<sup>5</sup> Si tratta di postazioni individuali dotate di strumentazione informatica per permettere, oltre alla ricerca catalografica, operazioni di lettura, di annotazione e di analisi dei testi e delle immagini. All'Archivo General de Indias, che raccoglie l'immensa documentazione riguardante l'impero coloniale spagnolo, è stato intrapreso, ormai da una decina d'anni, un grandioso programma di digitalizzazione completa dell'intero patrimonio documentario, allo scopo di favorire lo studio e l'accesso ai documenti e di proteggerli al tempo stesso dall'usura derivante dalla ripetuta consultazione diretta.<sup>6</sup> Questi due esempi sono paradigmatici per indicare la

<sup>5</sup> Cfr. J. VIRBEL, *Reading and Managing Texts on the 'Bibliothèque Nationale de France' Station*, in P. DELANY and G. LANDAU (eds.), *The Digital World. Text based computing in the humanities*, Cambridge Mass., The M.I.T. Press, 1993, pp. 31-52.

<sup>6</sup> Cfr. P. GONZALES, *The Digital Processing of Images in Archives and Libraries. Large*

direzione dello sviluppo. La portata di questa trasformazione è immensa. Indicandone la direzione, Manfred Thaller ha scritto un saggio dal titolo significativo: *L'archivio sul vostro tavolo di lavoro*.<sup>7</sup> È chiaro che tutto ciò implica, per la documentazione, un radicale aggiornamento teorico e qualitativo.

Con la rappresentazione integrale del documento, e con lo sviluppo della telematica, diventa possibile accedere, anche a distanza, non solo ad informazioni descrittive e analitiche *sul* documento (la rappresentazione catalografica o la documentazione bibliografica), ma alla rappresentazione integrale *del* documento e del suo contenuto. Questo fatto modifica sensibilmente i termini del rapporto tra documentazione e informatica. L'informatica non mette più le sue tecniche soltanto al servizio della gestione materiale del documento, o delle descrizioni catalografiche e bibliografiche che lo riguardano, ma le tecniche per il trattamento automatico dell'informazione vengono impiegate nella rappresentazione diretta e integrale del suo stesso contenuto. Diventa così possibile sostituire in misura sempre maggiore la rappresentazione digitale all'uso diretto del documento. In questo modo, come si è detto, la rappresentazione digitale contribuisce alla conservazione fisica del documento, evitandone il deterioramento. Ma è soprattutto la diffusione, la conservazione e l'integrità del *contenuto* del documento che viene a dipendere dalla qualità e dalla natura della rappresentazione digitale. La tecnologia dell'informazione, che è oggi in grado di produrre direttamente documenti elettronici, si trova a dover affrontare teoricamente il problema di fornire rappresentazioni adeguate dei documenti conservati nei grandi depositi storici, negli archivi e nelle biblioteche.

Così, il problema della rappresentazione digitale del documento storico diventa centrale per la formazione documentalistica dell'archivista e del bibliotecario. E la capacità di affrontare questo problema richiede, accanto alle competenze strettamente informatiche, competenze specifiche di natura filologica e storica. E non si tratta tanto di

*scale international projects*, in M. THALLER (ed.), *Images and Manuscripts in Historical Computing*, St. Katharinen, Max-Planck-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, 1992, pp. 97-121.

<sup>7</sup> Cf. M. THALLER, *The Archive on the Top of Your Desk? On Self-Documenting Image Files*, in J. FIKFAK and G. JARITZ (eds.), *Image Processing in History*, St. Katharinen, Max-Planck-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, 1993, pp. 21-44.

una pura e semplice giustapposizione, quanto di una precisa consapevolezza di natura epistemologica. Occorre la comprensione dell'impatto teorico che le procedure di elaborazione automatica possono avere sui procedimenti e sui metodi propri delle discipline umanistiche. In breve, è necessario l'apporto metodologico dell'informatica umanistica, la nuova disciplina di cui Tito Orlandi ha così bene delineato i confini.<sup>8</sup>

Anche a questo proposito saranno sufficienti alcuni esempi.

Si consideri il problema della rappresentazione del testo. Ora, come si legge in una trattazione specificamente dedicata al *text processing*, cioè all'elaborazione automatica dei dati testuali, «la parola 'testo' non è usata» dall'informatico «per significare il materiale letterario, così com'è stato prodotto direttamente dall'autore», ma per parlare di «informazione codificata in forma di caratteri, o di sequenze di caratteri».<sup>9</sup> La distanza tra le due concezioni è enorme. Da un lato il testo è fondamentalmente concepito come il *contenuto invariante* delle sue diverse rappresentazioni materiali, dall'altro è concepito come un *dato*, ossia come la rappresentazione codificata dell'informazione contenuta nel documento. La rappresentazione digitale non può quindi essere considerata come una semplice riproduzione *fisica* dei documenti, ma come una rappresentazione *logica*, costituita da dati elaborabili automaticamente. Può allora una pura e semplice «stringa di caratteri» riuscire a rappresentare tutta l'informazione contenuta nel testo? È stato così affrontato, e non da soli informatici, ma dalle associazioni per la linguistica computazionale (ACL) e le applicazioni letterarie e umanistiche dell'informatica (ALLC e ACH), il problema della codifica ulteriore dei dati testuali.<sup>10</sup> E ancora: può una struttura rigorosamente lineare del dato riuscire a rappresentare l'estrema complessità delle strutture testuali? Solo una rappresentazione in forma di *database* può riuscire, per esempio, a rappresentare un tipo di testualità 'fluida', come quello proprio di gran parte delle tradizioni testuali medievali. Problemi analoghi si pongono

<sup>8</sup> Cfr. T. ORLANDI, *Informatica umanistica*, Roma, La Nuova Italia Scientifica, 1990.

<sup>9</sup> A. C. DAY, *Text Processing*, Cambridge, Cambridge Univ. Press, 1984, pp. 1-2.

<sup>10</sup> Cfr. The Association for Computers and the Humanities (ACH), The Association for Computational Linguistics (ACL), The Association for Literary and Linguistic Computing (ALLC), *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, ed. by C. M. SPERBERG-MCQUEEN and L. BURNARD, Chicago-Oxford, TEI, 1994, 2 voll.

per le immagini digitali. Per scopi militari è stato sperimentato il riconoscimento automatico di oggetti ripresi con immagini satellitari. La presenza di nubi, però, anziché oscurare, poteva portare l'elaboratore al falso riconoscimento di carri armati, o di missili, cioè a vedere in esse proprio quello che esse dovevano celare. Anche l'immagine digitale è quindi un dato da analizzare logicamente e non una semplice riproduzione del suo oggetto. Si pone allora il problema della natura strutturale della rappresentazione: si tratta cioè di ottenere, con la rappresentazione digitale, strutture di dati adeguate al tipo di analisi che lo storico, o il filologo, ritengono di dover applicare al contenuto del documento.

Proprio allo scopo di affrontare in modo adeguato il problema della rappresentazione digitale e dell'elaborazione automatica dell'informazione trasmessa dai documenti storici, Manfred Thaller ha proposto e recato fondamentali contributi allo sviluppo di una rigorosa e metodologicamente autonoma *historical information science*,<sup>11</sup> una disciplina che si fonda sulla consapevolezza della natura affatto particolare dell'informazione trattata dallo storico — l'impossibilità di determinarne preliminarmente, o comunque in modo assoluto e definitivo il significato; una disciplina, inoltre, che viene di conseguenza guidata da due fondamentali criteri ispiratori: (1) trovare forme di rappresentazione potenzialmente esaustive e adeguate alla natura delle fonti; (2) mantenere separata la rappresentazione, il più possibile fedele, del documento dall'applicazione degli strumenti di analisi impiegati dallo storico per l'interpretazione del suo contenuto.

Se questo è il compito, la documentazione deve poterlo assolvere e la formazione deve tenerne conto: il conservatore, bibliotecario e archivistica, deve acquisire competenze documentalistiche che gli permettano di collaborare efficacemente con lo storico e col filologo alla produzione e all'uso di rappresentazioni digitali, scientificamente adeguate, delle fonti testuali e documentarie. In breve, e per concludere, il conservatore documentalista dev'essere dotato di competenze fornite non solo ed esclusivamente dall'informatica, ma anche e soprattutto dall'informatica umanistica.

<sup>11</sup> Cfr. M. THALLER, *What is 'Source-oriented Data Processing'? What is 'Historical Computer Science'?*, in L. I. BORODKIN and W. LEVERMANN (eds.), *Istoria i computer*, St. Katharinen, Max-Planck-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, 1993, pp. 5-18.