

La digitalizzazione retrospettiva dei periodici

Progetti e prospettive

di Antonella Farsetti

La retroconversione in formato digitale delle annate pregresse dei periodici sembra essere di crescente interesse per le biblioteche universitarie e nazionali. Per le prime, al fine di una migliore diffusione dell'informazione e del patrimonio in loro possesso, per le seconde anche in relazione alla diffusione e conservazione dell'eredità culturale di ogni nazione. Tutto ciò avrà sempre maggior importanza a livello internazionale e dovrà essere gestito nell'ambito di una pianificazione e cooperazione a livello nazionale ed allo stesso tempo europeo, per poter fornire un prodotto digitale esente da duplicazioni e spreco d'energie. Questo permetterà una fruizione condivisa, non solo dall'utente finale ma anche da parte delle istituzioni deentrici dell'importante "patrimonio culturale periodico" europeo. Questo lavoro vuole essere una breve panoramica su alcuni progetti considerati significativi dalla letteratura professionale. Senza pretese di esaustività, vorrei qui evidenziare alcune delle problematiche emerse all'interno di una parte di letteratura relativa agli ultimi due anni.

Parole d'ordine: coordinare e pianificare

In un clima di ferventi iniziative già implementate o in fase iniziale,¹ si stanno delineando linee guida gestionali che propongono modelli di coordinamento a livello nazionale all'interno di una cooperazione su scala europea. Possiamo fare qui due esempi: il primo è il JISC (Joint Information Science Council), che in ambito inglese ha coordinato e coordina numerosi progetti di digitalizzazione all'interno di eLib,² l'altro il progetto DEF (Denmark Electronic Research Library Programme),³ finan-

ziato dal Ministero della cultura, ricerca e istruzione danese.

La pianificazione a livello nazionale di DEF è stata resa possibile anche dal fatto che la Danimarca è una nazione relativamente piccola. Questo ha permesso al Ministero della cultura di finanziare il progetto con fondi elevati in proporzione alla quantità di materiale da recuperare.

Le realizzazioni più importanti

I progetti consolidati

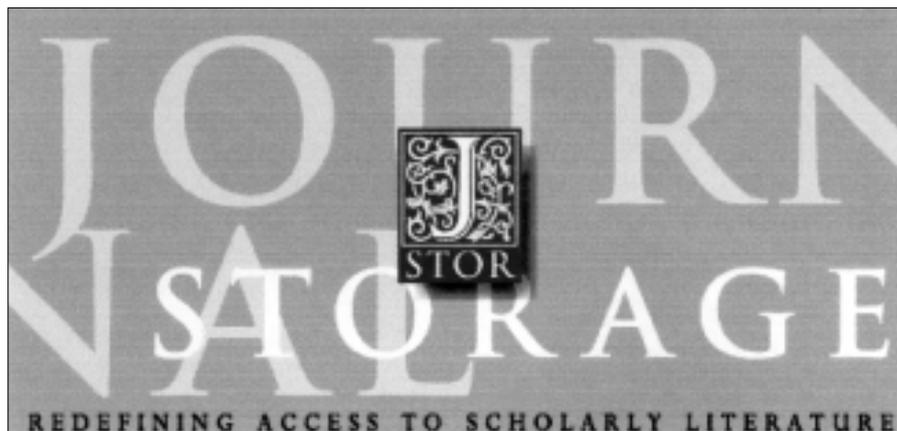
Quando parliamo di digitalizzazione di periodici, il progetto americano JSTOR⁴ è considerato nella maggior parte dei casi come punto di riferimento.

Tuttavia anche in Europa possiamo citare altre importanti esperienze.

Nel Regno Unito, durante gli ultimi quattro, cinque anni sono stati implementati vari progetti di digitalizzazione.

Tra questi ricordiamo Internet Library of Early Journals,⁵ CLIC (per *chemical communications*),⁶ "The journals and transactions of the Institute of British Geographers",⁷ il progetto BUILDER, nell'ambito del quale è stato digitalizzato "Midland History".⁸

Per la Francia possiamo citare GALLICA,⁹ ed infine per la Germania l'Università di Gottinga e il suo



centro di digitalizzazione, dove sono già stati implementati vari progetti relativi a collezioni monografiche rare e di pregio e dove è in corso il progetto DigiZeit.¹⁰

Progetti in fase di sviluppo

Questo nuovo progetto tedesco è finanziato dalla Deutsche Forschungsgemeinschaft (DFG) (German research foundation), che dal 1997 ha dato inizio ad un programma di finanziamento per la digitalizzazione retrospettiva di materiale librario.

DigiZeit, in linea con altri progetti europei, si propone di:

- fornire e facilitare l'accesso ai periodici più utilizzati e pubblicati da editori tedeschi;
- migliorare l'infrastruttura del ciclo informativo;
- migliorare la "visibilità" dei periodici scientifici tedeschi fornendo un accesso in linea.

Anche DigiZeit prenderà come modello l'americano JSTOR.

L'idea del modello JSTOR sembra incontrare anche il favore degli editori direttamente implicati nel progetto. Tale operazione può fornire agli editori stessi la versione elettronica delle loro pubblicazioni arretrate, senza grossi costi aggiuntivi.

L'importanza di una cooperazione a livello europeo è già evidenziabile negli intenti e nei primi risultati del progetto DIEPER.¹¹

DIEPER si prefigge di creare un unico accesso per i periodici digitalizzati in tutta Europa, in modo da permettere la ricerca su siti multipli. Sarà creato un database contenente informazioni relative a tutti i periodici che sono stati già retroconvertiti in formato digitale, in Europa e altrove, ed a quelli che sono in procinto di esserlo. Questo database potrà essere utilizzato per evitare la duplicazione dei titoli e per connettersi all'archivio digitale contenente il periodico già retroconvertito. Sarà inoltre possibile consultare direttamente il full-text

(in relazione alla politica di accesso adottata dai vari archivi correlati).

Gli standard

All'interno di questo contesto, proteso verso una cooperazione nazionale ed internazionale, si evince sempre più incisivamente la necessità di standard condivisi.

La mancanza di un coordinamento centrale, anche a livello europeo, può creare dissonanza tra i vari metodi di scansione, risoluzione e indicizzazione utilizzati.

I formati dovranno essere compatibili con gli sviluppi della tecnologia, poiché spesso si mira non solo ad un migliore accesso ai documenti ma anche alla loro conservazione. È necessario che gli standard siano aperti all'evoluzione tecnologica, perché se la digitalizzazione è interpretata anche come metodo di conservazione del patrimonio culturale dell'umanità, questa non può andare perduta o rischiare di non poter essere utilizzata dalle future generazioni per incompatibilità tra i software.

Nell'assenza di linee guida precise,¹² alcuni tra i maggiori progetti europei suggeriscono determinati standard in base alle loro esperienze.

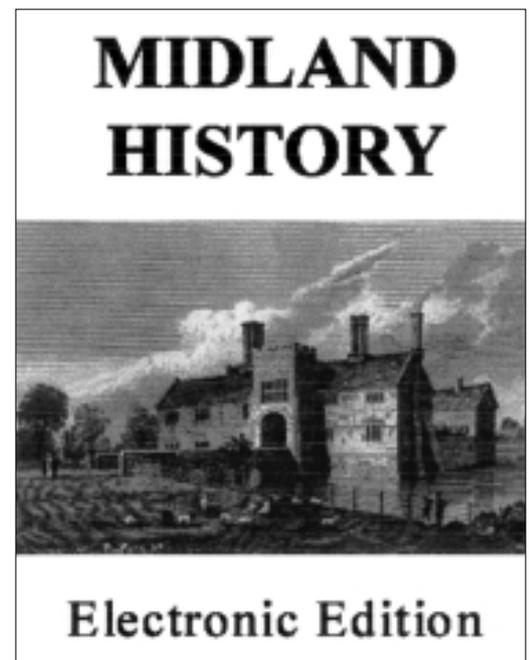
Questi standard sono spesso riconducibili a quelli suggeriti nell'ambito del Congresso di Copenaghen sulla digitalizzazione,¹³ tenutosi nel marzo 2000:

- scansione immagine a 600 dpi;
- metadata creati in XML/ RDF;
- full-text creato usando OCR in XML/TEI;
- unico identificatore per localizzare il full-text (URN o DOD).

La risoluzione dell'immagine

L'importanza della qualità del lavoro iniziale di scansione dell'immagine è decisamente elevata.

Ancor di più lo è quando si intende usare un software OCR per for-



nire la ricercabilità all'interno del testo.

L'uso di software OCR implica una serie di attività correlate il cui costo può avere notevole incidenza, sia da un punto di vista prettamente economico che da quello delle risorse umane.

Una scansione a 600 dpi è importante se si vuole poter applicare un software di information retrieval sul full-text.

Un alto livello di ricercabilità comporta un'alta quantità di correzioni. Tuttavia, migliore sarà l'immagine dalla quale partiamo, minori saranno gli interventi aggiuntivi necessari. Quando invece trattiamo volumi che non necessitino di ricerca ipertestuale, ma per i quali sia sufficiente fornire all'utente finale la loro immagine digitale (manoscritti, libri rari), uno standard di risoluzione 300/400 dpi è considerato sufficiente.

La risoluzione dell'immagine è comunque legata alle necessità dell'utente finale, che può essere soddisfatto da un'immagine con risoluzione non perfetta, ma che appaga comunque il suo bisogno informativo. ➤

È per questo che l'utente deve essere considerato il parametro fondamentale per la realizzazione di ogni progetto di digitalizzazione.

La selezione del materiale

Scegliere il materiale da digitalizzare non è un compito facile. Gli studiosi sono tentati di retroconvertire tutto. La difficoltà sta nel far capire loro che non bisogna decidere tra cosa gettare e cosa conservare; la scelta riguarda la selezione di quella parte di materiale da fare "emergere", rendendolo più accessibile all'utente.

L'interesse per i periodici digitali, relativi ad annate pregresse, non è solo appannaggio delle discipline umanistiche, ma coinvolge anche discipline scientifiche come la matematica e la chimica.¹⁴

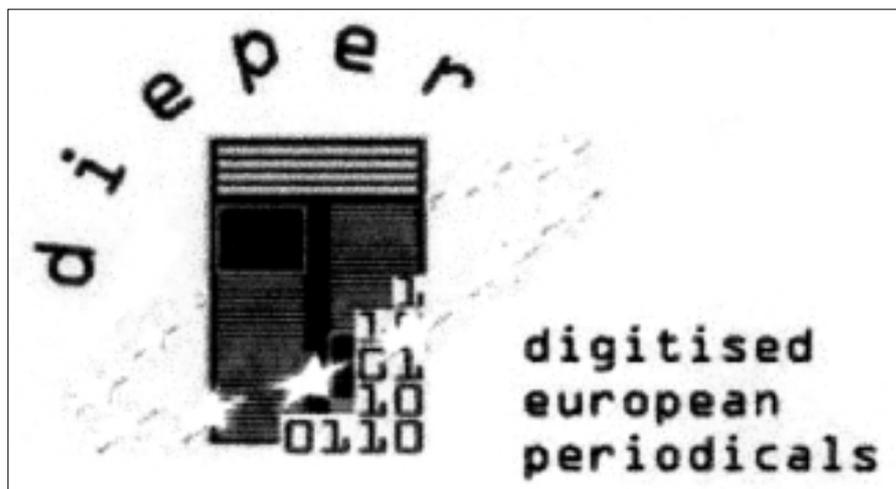
Il database del progetto DIEPER potrà essere un utile strumento di lavoro anche per la selezione del materiale, in modo da poter raggiungere criteri comuni.

La selezione del materiale, al momento, è raramente influenzata da una strategia nazionale e questo aumenta la difficoltà di una gestione europea.

Lo scenario attuale è molto variegato: si va da una situazione locale, dove i periodici vengono selezionati e digitalizzati con procedure *in-house*, scegliendo magari importanti riviste pubblicate anche all'interno della stessa struttura universitaria, ad uno scenario a larga scala come può essere quello del progetto BUILDER o il Glasgow digital library project.

La selezione dei periodici digitali all'interno di progetti ad ampio respiro, come quelli appena citati, è influenzata dalla necessità di spendere bene i soldi dei finanziatori, spesso rappresentati da sponsor esterni.

Le proposte di selezione sono spesso sottoposte a *peer review*. Ciò



sottintende anche un monitoraggio del materiale già pubblicato a livello internazionale.

I metadata

Creando informazione digitale non possiamo esimerci dall'affrontare le problematiche relative al suo reperimento. Dobbiamo quindi accennare anche all'uso che viene fatto dei metadata all'interno dei progetti di digitalizzazione.

I periodici sul web tendono ad essere meno stabili di quanto non lo siano quelli su carta. Non solo possono cambiare titolo, ma addirittura sparire.

Per il reperimento dell'informazione dei periodici in rete non ci possiamo limitare al livello di titolo, sul quale si basa l'ISSN, ma dobbiamo scendere al livello dell'articolo se vogliamo che la ricerca degli utenti sia agile e dia i risultati sperati.

Secondo Seamus Ross,¹⁵ nella digitalizzazione sono necessari tre tipi di metadata:

- a livello catalogafico;
- a livello tecnico, per visualizzare e conservare la risorsa;
- a livello amministrativo, per proteggerne i diritti.

I metadata inseriti nel testo vengono spesso indicati come alternativa

alla catalogazione tradizionale in relazione alle nuove esigenze di reperimento dell'informazione.

L'uso di questi metadata è efficace se la risorsa è strutturata ed è quindi possibile il loro inserimento all'interno del documento, sia esso effettuato da qual si voglia soggetto. Lo standard suggerito a livello europeo è XML/RDF per i metadata inseriti nel testo, poiché i documenti così strutturati sono ritenuti estremamente elastici nell'uso di applicativi di ricerca senza interventi aggiuntivi.



Stabilità e accesso

La grande sfida dei prossimi anni sarà mantenere le pubblicazioni elettroniche e permettere un continuo accesso ad esse.

Il mantenimento dell'accesso comporta costi ingenti, anche in relazione alla veloce obsolescenza del-

le attrezzature e dei software usati. In questo contesto si avverte nuovamente l'importanza dell'uso di standard condivisi e dei metadati per garantire un accesso a lungo termine.

I progetti CEDARS¹⁶ e NEDLIB¹⁷ sono due importanti iniziative in questa direzione.

Al fine di soddisfare i bisogni degli utenti sarà necessario porre particolare attenzione alle politiche di accesso ai documenti che dovranno essere aperte e compatibili con gli sviluppi che la tecnologia avrà nei prossimi anni.

Conclusioni

L'intero processo di digitalizzazione, dalla selezione alla definitiva implementazione del prodotto, in un contesto ideale dovrebbe essere suggerito da una strategia ad ampio respiro che esuli da isolate esigenze locali.

Molti sono gli interrogativi ai quali dovremo rispondere a breve, come gestori e creatori di informazione. L'augurio è che la velocità imposta dallo sviluppo delle nuove tecnologie si rifletta in una velocità di acquisizione di intenti comuni, e quindi di azione armonica, per il contesto italiano. ■

Note

¹ DigiZeit, <<http://www.deflink.dk>> e DIEPER <<http://www.sub.uni-goettingen.de/gdz/dieper/princip.htm>>.

² <<http://www.ukoln.ac.uk/elib/>>.

³ <<http://www.deflink.dk>>.

⁴ Journal Storage, <www.jstor.org/>.

⁵ <www.bodley.ox.ac.uk/ilej/>.

⁶ <www.ch.ic.ac.uk/clic/>.

⁷ <www.ingentajournals.bids.ac.uk/Pub-info/rgs.html>.

⁸ <www.builder.bham.ac.uk/midlandhistory/>.

⁹ <<http://gallica.bnf.fr>>.

¹⁰ <www.sub.uni-goettingen.de/gdz/de/main_en/html>.

¹¹ <<http://www.sub.uni-goettingen.de/>>.

[gdz/dieper/princip.htm](http://www.sub.uni-goettingen.de/gdz/dieper/princip.htm)>.

¹² SICCO, 1998.

¹³ *Digitising journals*, 2000.

¹⁴ EASON, 1998.

¹⁵ ROSS, 2000.

¹⁶ <<http://www.leeds.ac.uk/cedars/>>.

¹⁷ <<http://www.kb.nl/coop/nedlib/>>.

Riferimenti bibliografici

BAKKER, TRIX (1999), *Definition of collections, standards and procedures for retrospective digitisation*, "Liber quarterly", 9 (3), p. 305-322.

BORBINHA, JOSÉ LUIS (1998), *Digital libraries: a perspective from Portugal*, "Liber quarterly", 8 (1), p. 81-85.

BOURNE, ROSS (1998), *Towards bibliographic co-operation amongst European National Libraries*, "Liber quarterly", 8 (1), p. 106-109.

BREAKS, MICHAEL (1998), *The national electronic library: the united kingdom*, "Liber quarterly", 8 (1), p. 90-97;

CEDARS, <<http://www.leeds.ac.uk/cedars/>>.

CHILVERS, ALISON – FEATHER, JOHN (1998), *The management of digital data: a metadata approach*, "The electronic library", 16 (6), p. 365-371.

CLIC, <www.ch.ic.ac.uk/clic/>.

DE WITT, DONALD L., editor (1998), *Going digital: strategies for access, preservation and conversion of collections to a digital format*, New York, The Haworth Press, p. 233.

DEF Project, <<http://www.deflink.dk>>.

DIEPER, <<http://www.sub.uni-goettingen.de/gdz/dieper/princip.htm>>.

Digitising journals: conference on future strategies for european libraries, Copenhagen, 13-14 march 2000, <<http://www.deflink.dk>>.

DigiZeit, <http://www.sub.uni-goettingen.de/ebene_1/1_projek-e.htm>

Dublin Core Metadata Element Set, <<http://purl.org/dc/>>].

EASON, KEN [et al.] (1998), *A comparative analysis of the role of multimedia electronic journals in scholarly disciplines*, <<http://www.ukoln.ac.uk/services/elib/papers/tavistock/eason.htm>>.

Elib, <<http://www.ukoln.ac.uk/elib/>>.

FOOT, MIRJAM M., (1999), *Towards a preservation policy for european research libraries*, "Liber quarterly", 9 (3), p. 323-326.

FOX, PETER (1999), *Microfilming versus*

digitisation as a tool for preservation: long term access to digital material, "Liber quarterly", 8 (4), p. 436-447.

GALLICA Project, <<http://gallica.bnf.fr>>.

GUTHRIE, KEVIN M. (1999), *JSTOR: large scale digitisation of journals in the United States*, "Liber quarterly", 9 (3), p. 291-297.

HOLDEN, DANIEL (1999), *JSTOR: 1999 and beyond*, "Ariadne", <<http://www.ariadne.ac.uk/issue18/jstor/intro.html>>. Internet Library of Early Journals, <www.bodley.ox.ac.uk/ilej/>.

Journals and transactions of the Institute of British Geographers (The), <www.ingentajournals.bids.ac.uk/Pub-info/rgs.html>.

JSTOR, <<http://www.jstor.org/about/board.html>>.

KLUGKIST, ALEX C. (1998), *Digital information, licensing and consortia*, "Liber quarterly", 8 (1), p. 110-112.

Licensing principles: guidelines and checklist for libraries, "Liber quarterly", 8 (1), p. 113-117.

Midland History, <www.builder.bham.ac.uk/midlandhistory/>.

NEDLIB, <<http://www.kb.nl/coop/nedlib/>>.

NOORDERMEER, TRUDI [et al.] (1998), *Electronic library developments in the Netherlands*, "Liber quarterly", 8 (1), p. 57-80.

PRESIDENZA DEL CONSIGLIO DEI MINISTRI, *A new european social agenda: a proposal for action from the british and italian governments*, International society forum newsletters, Documents (1999), <http://www.palazzo.chigi.it/fsi/eng/info/news_eng.html>.

RONCAGLIA, GINO (2000), *L'informazione diventa digitale*, p. 5-33, in CIOTTI, FABIO – RONCAGLIA, GINO, *Il mondo digitale: introduzione ai nuovi media*, Bari, Laterza, p. 510.

ROSS, SEAMUS (2000), *The success and failure of digitisation projects: models for building digital libraries*, [Conferenza, Roma, 24 Ottobre 2000, non pubblicata].

SCHWARTZ, WERNER (1999), *The DIEPER project, Digitised European PERiodicals*, "Liber quarterly", 9 (3), p. 298-304.

SCOAP Project, <<http://www.nla.gov.au/scoap/>>.

SICCO, MARIA (1998), *Guidelines for digitizing collections in italian libraries*, "Liber quarterly", 8 (1), p. 49-56.