

Biblioteche digitali e studi umanistici

Maurizio Lana¹

¹ Dipartimento di Studi Umanistici/Università del Piemonte Orientale, Vercelli, Italia
m.lana@lett.unipmn.it

Abstract. Il complesso e costoso lavoro di creazione di biblioteche digitali non si può giustificare solo sulla base della semplice creazione di una risorsa per la lettura dei testi nell'ambito digitale. Occorre che le biblioteche digitali siano utilizzate per attività complesse non praticabili con le edizioni a stampa. Una di esse è l'annotazione formale dei testi e l'impiego dell'annotazione per dar luogo a nuove modalità di lavoro sul testo stesso.

Parole chiave: geografia, ontologia, TEI, Linked Open Data.

1. Introduzione

Accade in genere che davanti ad un fenomeno nuovo si provi in primo luogo a leggerlo in analogia con qualcosa di preesistente che appare simile. Fotografia digitale e fotografia tradizionale (analogica) sono diversissime, le loro somiglianze in fin dei conti sono marginali. Però si parla sempre di fotografia, stampe, esposizione, fotocamera, e così via. La diffusione è facilitata dal fatto che ciò che nella forma tradizionale appariva difficile – fare buone foto – sembra ora facile con tutte le preimpostazioni disponibili per cui le foto sbiancate per la sovraesposizione o nere di sottoesposizione non esistono più. Così facendo però si perde di vista che la fotografia digitale permette un controllo diretto sulla produzione dell'immagine più ampio e diretto di quello che era possibile con la pellicola. E che permette anche operazioni non possibili con la pellicola.

Con le biblioteche digitali si rischia che accada qualcosa di simile. Costruire una biblioteca digitale, soprattutto se si tratta di una biblioteca costruita a partire da fonti a stampa digitalizzate e poi accuratamente corrette e annotate, richiede investimenti cospicui in tempo e denaro. E nel corso degli anni occorre aggiornare il formato dei dati, il software di gestione delle biblioteche, e le opere digitalizzate sono meno leggibili che se si trovassero su un supporto fisico: la conservazione è delicata e difficile – non basta 'tenere i libri lì', occorre curarli e accudirli costantemente come bottiglie di champagne nella *cave*.¹

¹ Più obbligato il percorso per le biblioteche digitali che danno accesso a collezioni nativamente digitali: rivi-

Di qui la domanda: perché costruire biblioteche digitali? La domanda ovviamente si pone essenzialmente per le biblioteche le cui collezioni sono state digitalizzate, perché quando esse invece custodiscono collezioni *born digital* la risposta in certo modo è apparentemente data dal fatto stesso che esistono contenuti che nascono digitali – libri, giornali, film, musiche, giochi e dunque richiedono una biblioteca digitale che li raccolga e li renda disponibili. La risposta è che la digitalizzazione deve essere il ponte, la rampa di lancio, verso utilizzi, attività, impossibili o comunque non facili da realizzare se i contenuti si trovano ancorati ad un supporto fisico.

Dunque – provocatoriamente e paradossalmente – una biblioteca che contenga collezioni *digitalizzate*² non ha come scopo primario di offrire l'accesso in lettura ai libri che possiede: le edizioni a stampa già esistenti sono più facili da gestire, distribuire, leggere. In misura più o meno grande, tutti i modi di accesso ai contenuti digitali che ripropongano *primariamente* le forme consuete per l'accesso ai corrispondenti contenuti su supporto fisico costituiscono uno spreco di risorse. Il nucleo della questione sta nell'annotazione: cioè nello scrivere in modo formalizzato informazioni associate ad uno specifico passo di testo. Nulla di nuovo: è 'solo' la versione tecnologicamente evoluta dell'annotazione a margine che qualsiasi lettore esperto è abituato a fare, e che ha la sua forma più illustre negli *scholia* che ci testimoniano l'attività di commentatori medievali delle opere della letteratura greco-latina. Non solo: il testo, soprattutto letterario (ma in realtà tutti, in vario modo) vive solo nella lettura di un soggetto che con quell'atto lo attiva. Perché il lettore interpreta ciò che legge, individua significati, costruisce visioni delle cose, a partire dal testo letto. Legge con un'intenzionalità, con uno scopo grazie ai quali acquisisce conoscenza, individua informazioni, formula ragionamenti, relativi a specifici punti o passi del testo. Tutto ciò si può tradurre in annotazioni formalmente rigorose che possono essere riusate e rielaborate: e ciò è precisamente uno dei modi in cui si manifesta lo scopo proprio di una biblioteca digitale, non semplicemente leggere i testi, ma lavorare su di, e con, essi in modi altrimenti (praticamente) impossibili.

Il percorso quindi è segnato da queste tappe:

testo a stampa → digitalizzazione → testo digitale → lettura/correzione →
annotazione → rielaborazione del testo annotato.

² Proprio solo per semplicità espositiva parleremo nelle pagine seguenti di *libri* come contenuto delle biblioteche digitalizzate, ma essi sono linguisticamente solo un modo semplice per indicare "i contenuti digitalizzati, di qualsiasi tipo".

Nelle pagine che seguono si descriverà una biblioteca digitale in corso di realizzazione e si mostrerà come la sua esistenza permetta di costruire un progetto di ricerca altrimenti irrealizzabile.

2. La biblioteca digitale digilibLT

Gli studiosi del mondo classico hanno a disposizione due raccolte di testi su CDROM realizzate negli anni 90 del secolo scorso, note come TLG (Thesaurus Linguae Graecae) e PHI (Packard Humanities Institute) CDROM. Mentre il primo raccoglie tutti i testi in lingua greca antica dalle origini all'epoca bizantina, esaurendo così l'orizzonte temporale di quella che viene chiamata "letteratura greca" sia nella componente classica sia in quella bizantina, il PHI CDROM raccoglie testi in lingua latina dalle origini all'epoca classica con ciò intendendo testi che si collocano intorno al I secolo dopo Cristo. Rimane scoperto quindi il periodo che va dal I/II secolo dopo Cristo fino alla caduta dell'impero romano d'Occidente nel 456 d.C. cioè il periodo del latino tardo (o tardoantico). Di qui l'idea di colmare la lacuna sia per amore di completezza sia perché i testi latini tardi sono quelli che hanno trasmesso il pensiero e la civiltà latina al Medioevo³.

Poiché le opere latine tarde sono tutte disponibili in edizioni a stampa, oppure in edizioni online costose, oppure in edizioni online di non eccelsa qualità e/o sparse e/o dallo statuto formale non sempre chiaro, creare una biblioteca digitale specialistica che le accogliesse significava digitalizzare il testo stabilito delle edizioni a stampa, con l'intento di costruire una risorsa di alto livello non solo per i latinisti ma anche più in generale per tutti coloro che studiano il mondo antico. Prima questione da affrontare, la necessità di fondi mirati alla realizzazione della biblioteca. Nel 2008 in risposta ad un Bando della Regione Piemonte che finanziava iniziative nell'ambito delle scienze umane e sociali⁴ venne dunque concepito e redatto da chi scrive un progetto che fu presentato, e dopo aver attraversato un processo di *blind peer evaluation* nell'agosto 2009 fu approvato e cofinanziato. La biblioteca attualmente è in corso di realizzazione presso il Dipartimento di Studi Umanistici dell'Università del Piemonte Orientale ed è disponibile all'indirizzo <http://www.digiliblt.unipmn.it>. Il progetto non nasceva da bibliotecari ma da studiosi interessati ad uno specifico ambito disciplinare i quali poi avevano coinvolto un bibliotecario nella progettazione⁵.

³ Idea di Raffaella Tabacco, latinista del Dipartimento di Studi Umanistici dell'Università del Piemonte Orientale.

⁴ <http://www.regione.piemonte.it/innovazione/ricerca/bandi-e-finanziamenti/bandi-aperti/bando-scienze-umane-e-sociali.html> [ultima visita 28.5.2014].

⁵ La direttrice della biblioteca del Dipartimento, Silvia Botto.

Molti aspetti si potrebbero qui ricordare ma vale la pena soffermarsi su alcuni in particolare in relazione allo spazio disponibile. In primo luogo il tema della *openness*: il sito della biblioteca è costruito interamente con software *open source*⁶ e tutto il suo contenuto è disponibile in *open access*. Il tema immediatamente sotteso a queste scelte è che ciò che viene realizzato con fondi pubblici deve essere liberamente accessibile al pubblico che tramite il pagamento delle tasse ne ha resa possibile la realizzazione; ma nel contempo deve anche costare il meno possibile – senza detrimento per la qualità sia specifica delle varie componenti, sia complessiva – così da massimizzarne l'efficacia; e deve per quanto possibile durare nel tempo con costi vivi⁷ ridotti al minimo. Ma è almeno altrettanto rilevante, se non di più, il fatto che una biblioteca basata su un sistema aperto è più facilmente interoperabile (cioè predisposta per poter scambiare dati con altre biblioteche): i formati dei dati sono aperti e pubblici anch'essi⁸, le librerie di software necessarie per l'interazione sono pubbliche, i programmi stessi sono configurabili e modificabili secondo le necessità dell'utilizzatore. Naturalmente ci si potrebbe chiedere perché mai si debba ritenere importante l'interoperabilità. La ragione è che in generale i dati – e il contenuto di una biblioteca digitale è costituito da dati – sono tanto più interessanti quanto più sono numerosi e vari; e per aumentarne la numerosità e varietà il modo più semplice e valido è quello di riunire dati provenienti da differenti origini. Nello specifico: 'riunire' libri digitali provenienti da differenti biblioteche (che cosa esattamente significhi 'riunire' dipende di volta in volta dalle intenzioni e dalle modalità operative scelte⁹) le quali devono essere interoperabili perché questa unificazione dei dati possa avvenire.

Poi c'è il tema della qualità dei testi e del loro trattamento digitale. E' interessante un'osservazione formulata in ambito storico ma ugualmente valida anche in ambito letterario da Tim Hitchcock, condirettore di Old Bailey Online, che ha scritto:

This discussion piece argues that the design and structure of online historical resources and the process of search and discover embodied within them create a series of substantial problems for historians.

⁶ Il sistema operativo è Ubuntu Server, il server web Apache, il linguaggio di programmazione PHP, il library manager è XTF, il database MySQL.

⁷ Quali lo spazio sul server, la corrente elettrica di alimentazione, un servizio di backup, e simili, che consentono l'esistenza del servizio offerto. Non costi di licenze di software, ad esempio.

⁸ E' noto per concetto e per esperienza che formati proprietari per i dati, derivanti dall'utilizzo di un software commerciale, possono rendere impossibile l'interazione profonda con altre biblioteche.

⁹ Due principalmente sono le opzioni: costruire un'interfaccia di consultazione e ricerca comune a differenti biblioteche e quando l'utente formula una ricerca ridistribuirla alle funzioni di ricerca delle varie biblioteche aderenti; oppure raccogliere centralmente i libri digitali delle varie biblioteche, e lì effettuare le ricerche richieste dall'utente; in entrambi i casi quando l'utente clicca su un esito viene indirizzato al sito della biblioteca pertinente.

Algorithm-driven discovery and misleading forms of search, poor OCR, and all the selection biases of a new edition of the Western print archive have changed how we research the past, and the underlying character of the object of study. (Hitchcock 2013)

L'aspetto importante qui è la notazione "poor OCR", Hitchcock afferma che quando si digitalizzano fonti testuali a stampa, la scarsa qualità del riconoscimento ottico del testo crea una serie di problemi sostanziali agli studiosi. Consapevoli di questo in digilibLT ogni testo digitalizzato viene (ri)letto almeno tre volte: le prime due sono vere e proprie correzioni di bozze¹⁰, la terza è la fase di annotazione formale del testo.

Un terzo aspetto rilevante è quello della licenza d'uso dei contenuti. Distribuire i contenuti in una prospettiva di *open access* non significa abbandonarli nel web come un tronco portato dal mare su una spiaggia; significa invece assegnare ad essi in modo esplicito e formale una licenza d'uso che indichi in quali modi se ne consente la redistribuzione e circolazione in forme diverse da quelle del diritto d'autore tradizionale. C'è oggi una spinta crescente da parte per esempio del programma *Horizon 2020* e della *Budapest Open Access Initiative 2012* (giusto per citare due soggetti importanti in questo ambito) verso l'adozione di licenze Creative Commons molto aperte, basate sulle clausole BY (attribuzione) e SA (condividi allo stesso modo) orientate a permettere il riuso facile e privo di vincoli economici dei contenuti così distribuiti, su cui la discussione è aperta¹¹.

Ultimo, tra i tanti spunti possibili da approfondire, la biblioteca si sta espandendo oltre i confini del latino letterario tardoantico: da un lato con l'acquisizione degli scritti noti come "grammatici latini": un insieme di opere di argomento linguistico/grammaticale di autori che si collocano tra il II e il V/VI sec. d.C.; dall'altro con l'acquisizione delle opere già raccolte nel PHI CDROM perché su tale collezione di testi i diritti di proprietà intellettuale sono scaduti e dunque essa può essere riutilizzata. L'idea guida è che l'esistenza di molteplici iniziative che mirano a costruire una biblioteca digitale globale del latino¹² è positiva in quanto ciascuna di essa potrà offrire ai lettori edizioni differenti delle medesime opere. E l'interoperabilità, in questo caso, significherebbe che lo studioso può cercare

¹⁰ Accade talora che in questa rilettura accurata vengano individuati veri e propri errori tipografici presenti nell'edizione a stampa.

¹¹ Non è questa la sede, e non c'è lo spazio, per discutere approfonditamente di queste licenze che sono descritte nel sito Creative Commons all'indirizzo <http://www.creativecommons.it/> [ultima visita 28.5.2014]. Sulla complessa questione dei maggiori o minori vincoli all'utilizzo a fini di lucro dei contenuti distribuiti con licenze Creative Commons, si veda (Lana 2014).

¹² Si possono citare almeno l'iniziativa dell'American Philological Association – che non ha ancora dato luogo ad una biblioteca digitale esistente; e quella dell'Open Philology Project di Gregory Crane Humboldt Professor a Lipsia, <http://www.dh.uni-leipzig.de/wo/projects/> [ultima visita 28.5.2014]. Senza dimenticare, per la sua ampiezza, la biblioteca disponibile in Perseus, <http://www.perseus.tufts.edu/> [ultima visita 28.5.2014].

l'opera di un autore e vedere in quali edizioni essa è disponibile nelle varie biblioteche digitali.

3. Il progetto geolat

Avendo a disposizione una biblioteca digitale di testi latini si può pensare a modalità di studio e lavoro sui testi prima non praticabili. Il progetto geolat (<http://www.geolat.it>)¹³ lavora per costruire nuovi modi di accesso ai testi latini di digilibLT valorizzandone i contenuti geografici: per mezzo di un'ontologia geografica del mondo antico appositamente costruita, sarà possibile annotare formalmente i nomi geografici presenti nei testi utilizzando lo standard TEI e pubblicare le annotazioni in forma di Linked Open Data. Sulla base di tali annotazioni verranno prodotte carte geografiche arricchite dall'accesso ai testi della biblioteca: tracciando un'area sulla carta si otterrà l'elenco degli autori, opere, e passi, in cui sono presenti nomi geografici appartenenti a quell'area; scelto un luogo sarà possibile individuare sia in forma testuale sia in forma cartografica i luoghi che cooccorrono con il primo; oppure, scelto un luogo leggere i passi dei testi che lo menzionano, e così via.

Anche geolat si basa sui medesimi principi esposti per la biblioteca digilibLT, cioè *open access*, licenze Creative Commons, software open source; e l'utilizzo del meccanismo dei Linked Open Data permetterà l'interoperabilità con altre risorse riguardanti il mondo antico, prima fra tutte il gazzettiere Pleiades¹⁴: da Pleiades geolat prenderà per ogni nome di luogo le coordinate geografiche e la traduzione del nome latino nelle lingue contemporanee in cui esso è disponibile; mentre Pleiades potrà prendere da geolat per ogni nome di luogo l'elenco dei passi in cui esso ricorre.

In questo modo due differenti risorse si arricchiscono reciprocamente e ciò è reso possibile dalla disponibilità di una biblioteca digitale.

4. Bibliografia

Talbert R., Bagnall R., a c. di (2000). *Barrington Atlas of Greek and Roman World*, Princeton University Press, Princeton

¹³ Finanziato dal 2012 al 2015 dalla Compagnia di San Paolo a seguito di una blind peer evaluation effettuata da European Science Foundation. [ultima visita 28.5.2014]

¹⁴ <http://pleiades.stoa.org/> [ultima visita 28.5.2014]. La fonte di autorità di Pleiades è (Talbert, 2000).

Lana M. (2014). Licenze d'uso e valorizzazione della ricerca umanistica. «Digitalia», n. 1. URL=<http://digitalia.sbn.it/> [in stampa].

Hitchcock T. (2013). Confronting the Digital: or How Academic History Writing Lost the Plot. «Cultural and Social History», vol. 10, n. 1, pp. 9-23.

1.1 Riferimenti bibliografici

I riferimenti vanno riportati nel corpo del documento e non in nota a piede pagina utilizzando lo stile di citazione "APA": fra parentesi tonde con cognome dell'autore anno, pagina o intervallo di pagine (es. Buzzetti 1998, 34).

Nella bibliografia finale, che si chiamerà Bibliografia e sarà formattata come un titolo di primo livello numerato progressivamente, i riferimenti andranno riportati con sporgenza di prima riga di 1,25. I riferimenti andranno sciolti nelle seguenti forme:

Monografie. Cognome dell'autore Iniziale puntata del nome. (anno). Titolo in corsivo, Editore.

Es.

- Goldfarb C.F. (1990). *The SGML Handbook*, Oxford University Press.
- McGann J. (2002). *La letteratura dopo il World Wide Web. Il testo letterario nell'era digitale*, a c. di D. Buzzetti, Bononia University Press.
- Burnard L., O'Brien O'Keefe K., Unsworth J., a c. di (2006). *Electronic Textual Editing*, MLA. URL=<http://www.tei-c.org/Activities/ETE/Preview/index.xml> [ultima visita 18.12.2012].

Contributo in volume. Cognome dell'autore Iniziale puntata del nome. (anno). Titolo in corsivo. In Iniziale puntata del nome Cognome dell'autore, Titolo del volume, Editore, p. o pp.

Es.

- Hoover D. (2008). *Quantitative Analysis and Literary Studies*. In S. Schreibman, R. Siemens, J. Unsworth, a c. di, *A companion to Digital Literary Studies*, Blackwell, pp. 218-256. URL=<http://www.digitalhumanities.org/companionDLS/> [ultima visita 8.12.2012].

Articolo in rivista. Cognome dell'autore Iniziale puntata del nome. (anno). Titolo in corsivo. «Nome rivista», vol., no, p. o pp.

Es.

- Bradley J. (2008). *Pliny: A model for digital support of scholarship*. «Journal of Digital Information», vol. 9, no 1. URL=<http://journals.tdl.org/jodi/index.php/jodi/article/view/209/198>. [ultima visita 8.2.2012].

Atti di convegno. Cognome dell'autore Iniziale puntata del nome. (anno), Titolo in corsivo. In Titolo del convegno per esteso, Editore, p. o pp.

Es.

- Marshall C.C. (1998). *Towards an Ecology of Hypertext Annotation*. In Proceedings of Hypertext 98, ACM, pp. 40-49.
- Agosti M. *et al.* (2011). *A Digital Library of Grammatical Resources for European Dialects*. In *Digital Libraries and Archives*, 7th Italian Research Conference, IRCDL 2011 Revised Papers. Communications in Computer and Information Science, vol. 249, Springer-Verlag, pp. 61-74.

Norme generali. Norme da utilizzare in ogni caso occorra una delle seguenti casistiche:

- Se c'è una curatela utilizzare la formula a c. di Iniziale puntata del nome Cognome.
- Se gli autori sono fino a 3 vanno inseriti separandoli da una virgola. Se sono più di tre basta il primo e la menzione *et al.*
- Se è necessario citare un URL va in forma estesa, con indicazione di autore, titolo e data (se presente) - nella forma da utilizzare per ogni tipologia di contributo - indirizzo Web [data di ultima visita].
- Se un contributo è anche online usare la forma URL=indirizzo completo di protocollo.

1.2 Abbreviazioni

Da utilizzarsi sia nel corpo del documento che nella bibliografia. Attenersi alle forme sotto indicate (la lista non è esaustiva). Di norma, il plurale si ottiene raddoppiando l'ultima consonante (ess.: artt. = articoli; pp. = pagine):

app. = appendice	f. = foglio	p. = pagina
art. = articolo	fasc. = fascicolo	r = recto
c. = carta	fig. = figura	r. = riga
ca. = circa	ibid. = ibidem	s. = seguente
cap. = capitolo	id. // ead. = idem, eadem	s.a. = sine anno
cfr. = confronta	ill. = illustrazione	s.d. = senza data
cit. = citato	it. = italiano (-a)	sec. = secolo
cod. = codex // co- dice	l. = linea	sez. = sezione
col. = colonna	l.c. = luogo citato	s.l. = sine loco
dx. = destro (-a)	misc. = miscella- neo (-a)	s.t. = sine ty- pographo
ed. = edizione	ms. = manoscritto	suppl. = supple- mento
e.g. = exempli gra- tia	n. = nota	sx. = sinistro (-a)
es. = esempio	n.s. = nuova serie	t. = tomo
etc. = eccetera,	no = numero	tab. = tabella

tav. = tavola	v. = verso [di un componimento poetico]	[...] = verba ab editore ommissa
trad. = traduzione		
v = verso [di una carta o di un foglio]	v.l. = varia lectio	
	vol. = volume	

2 Bibliografia

- Agosti M. *et al.* (2011). *A Digital Library of Grammatical Resources for European Dialects*. In *Digital Libraries and Archives*, 7th Italian Research Conference, IRCDL 2011 Revised Papers. Communications in Computer and Information Science, vol. 249, Springer-Verlag, pp. 61-74.
- Bradley J. (2008). *Pliny: A model for digital support of scholarship*. «Journal of Digital Information», vol. 9, no 1. URL=<http://journals.tdl.org/jodi/index.php/jodi/article/view/209/198>. [ultima visita 8.12.2012].
- Burnard L., O'Brien O'Keeffe K., Unsworth J., a c. di (2006). *Electronic Textual Editing*, MLA. URL=<http://www.tei-c.org/Activities/ETE/Preview/index.xml> [ultima visita 18.12.2012].
- Goldfarb C.F. (1990). *The SGML Handbook*, Oxford University Press.
- Hoover D. (2008). *Quantitative Analysis and Literary Studies*. In S. Schreibman, R. Siemens, J. Unsworth, a c. di, *A companion to Digital Literary Studies*, Blackwell, pp. 218-256. URL=<http://www.digitalhumanities.org/companionDLS/> [ultima visita 8.12.2012].
- Marshall C.C. (1998). *Towards an Ecology of Hypertext Annotation*. In Proceedings of Hypertext 98, ACM, pp. 40-49.
- McGann J. (2002). *La letteratura dopo il World Wide Web. Il testo letterario nell'era digitale*, a c. di D. Buzzetti, Bononia University Press.