

Dig *Italia*

Anno VII, Numero 1 - **2012**

Rivista del digitale nei beni culturali

ICCU-ROMA

Metodologie e problematiche per una biblioteca digitale. Il caso di digilibLT

Maurizio Lana

Università del Piemonte Orientale, Dipartimento di studi umanistici

Una biblioteca digitale, oggi, non è più in se stessa un prodotto altamente innovativo, ma uno strumento consueto di lavoro e dunque deve per quanto possibile rispondere nel modo migliore a una molteplicità di esigenze e vincoli.

digilibLT (digital library of late-antique latin texts; <http://digiliblt.unipmn.it>) è una biblioteca digitale nata di recente con lo scopo di raccogliere tutti i testi latini pagani tardi, a completamento della raccolta dei testi latini di epoca classica pubblicata nel PHI CDROM #5.3. La biblioteca è stata progettata in risposta ad un bando della Regione Piemonte, da cui è cofinanziata, con l'intento di rispondere alle esigenze più avvertite tra gli studiosi e i bibliotecari. I testi antichi, che sono/saranno disponibili per il download in molteplici formati, da XML-TEI a testo, PDF ed epub (oltre al 'vecchio formato' Beta della raccolta PHI 5.3, per ragioni di compatibilità verso il passato), per rispondere alle esigenze di qualità degli studiosi provengono tutti da edizioni critiche recenti e selezionate dal comitato scientifico della biblioteca; introduzioni, apparati e commenti presenti nelle medesime edizioni sono invece del tutto assenti, in ossequio alla legge sul diritto d'autore. Il sistema è costruito interamente con free software, adottando senza limitazioni i principi dell'Open Access e le licenze Creative Commons per tutti i contenuti (pagine Web e testi).

Una biblioteca digitale oggi

Una biblioteca digitale si può presentare con molte facce, in relazione al tipo di testi che raccoglie e alla finalità con cui li raccoglie: da testi scientifici pubblicati su riviste professionali e già esistenti in formato digitale, fino a testi letterari pubblicati in edizioni critiche a stampa che devono essere digitalizzati. La biblioteca di cui si parlerà nelle pagine seguenti è del secondo tipo (si parte da testi letterari a stampa); quindi anche se non ci sono questioni da risolvere legate all'interazione con editori di riviste digitali (prima fra tutte la contrattazione per la ridistribuzione dei contenuti) rimane necessario operare una serie di scelte di tipo sistemico sul genere di accesso da fornire ai lettori e sui vincoli da porre a tale accesso. Non più di 15 anni fa una raccolta di testi su CD-ROM costituiva una risorsa di studio importante, preziosa, tecnicamente avanzata, oggi non più. Oggi ad una raccolta di testi digitali chiediamo di essere una biblioteca digitale. Cioè non solo di fornire i testi ma di collocarli dentro un sistema che li descriva, li valorizzi, li metta in relazione con altri testi e con altre biblioteche che li possiedono. In altre parole intorno ai testi cerchiamo dei servizi.

Come è nato il progetto digilibLT

Già una decina d'anni fa, all'interno del Dipartimento di Studi Umanistici dell'Università del Piemonte Orientale, Raffaella Tabacco, docente di Letteratura latina, aveva iniziato a pensare insieme con chi scrive a una biblioteca digitale del latino tardo (dalla fine del I secolo d.c. fino alla fine del V secolo d.c.), avendo presente come riferimento quello che gli studiosi di testi classici chiamano "CD-ROM del PHI" cioè il *PHI CD-ROM #5.3* che il Packard Humanities Institute aveva pubblicato nel 1991 contenente i testi latini dalle origini fino alla fine del I secolo d.c./inizio del II. Inizialmente, negli anni Ottanta del secolo scorso, la piattaforma di riferimento per le banche dati di testi greci e latini su CD-ROM era stata il mini-computer Ibycus (fig. 1) prodotto da Hewlett Packard.



Figura 1. Computer Ibycus (<http://www.tlg.uci.edu/about/ibycus.php>)

L'idea del progetto digilibLT era di colmare la lacuna data dal fatto che la base di dati testuale contenuta nel CD-ROM del PHI non ebbe un completamento che la portasse a raccogliere tutta la letteratura latina (come invece è il caso della letteratura greca classica e del TLG CD-ROM), e ciò si voleva fare sia per necessità di completezza (se si può dire, un po' ovviamente, un po' convenzionalmente, che i testi latini vanno

dalle origini di Roma fino alla sua fine¹, allora una raccolta digitalizzata dei testi latini² deve comprendere tutti quei testi), sia perché la cultura tardo antica ebbe un

¹ La data convenzionale è il 476 d.C. ma gli scrittori latini tardi si collocano intorno ad essa e magari anche dopo.

² Il termine *raccolta* in parte è ambiguo: perché se si ragiona dal punto di vista fisico, del luogo cioè dove sono raccolti i testi, una prima metà di essi si trova nel CD-ROM dei testi latini del PHI, mentre la seconda metà a cura del progetto digilibLT ovviamente non potrà essere collocata sul medesimo supporto; e quindi, dal punto di vista fisico dei supporti, si tratterà sempre di più di una raccolta. Ma dal punto di vista strutturale, cioè del tipo di formato dei file, anche i testi latini tardi saranno codificati come già quelli del PHI (su questo si vedano più sotto le pagine dedicate al tema della codifica (p. 9), nelle due forme TEI (p. 11) e (Beta code (p. 10)), e quindi l'unità della

ruolo importante nella trasmissione all'Europa moderna del pensiero del mondo antico in materia di scienza, politica, filosofia, letteratura, religione e in assenza di una raccolta completa di testi digitali è difficile se non impossibile effettuare studi su tale letteratura ricorrendo ai metodi ormai correnti per lo studio di testi di epoca classica. Non si tratta di un'esigenza d'avanguardia o di un piccolo gruppo di superesperti, perché il mondo degli studiosi di testi, e di testi classici, è stato particolarmente attivo negli impieghi di ricerca dell'informatica fin dagli anni Quaranta (senza pretesa di completezza si può ricordare che Roberto Busa, iniziatore delle concordanze prodotte per mezzo dei computer, avviò la sua attività con l'IBM nel 1949; datano al 1984 il volume *Computer programs for literary analysis* di John R. Abercrombie, pubblicato da University of Pennsylvania Press; e la newsletter «OFFLINE», diretta da Robert Kraft)³.

Le biblioteche digitali per il latino non mancano, ma molte di esse, o distribuite su CD-ROM (CETEDOC Library of Christian Latin Texts, Bibliotheca Teubneriana Latina, CSEL, per esempio), talora con problemi di disponibilità, o accessibili sul Web (Patrologia Latina Database e Bibliotheca Teubneriana Latina per citarne un paio) sono proprietarie, cosa che implica costi di accesso sovente molto alti e limitazioni agli usi possibili, la più seria essendo costituita dal fatto che le ricerche testuali si fanno solo con un programma di analisi fornito dall'editore dei testi; seguita dalla difficoltà o impossibilità di salvare localmente i testi nella loro interezza. Testi latini isolati si trovano sparsi in rete nelle pagine Web di istituzioni, studiosi, appassionati. I limiti in questo caso sono costituiti di solito dall'assenza di dichiarazioni esplicite riguardanti la condizione di correttezza del testo, l'edizione di riferimento, le condizioni d'uso, e il fatto che sono assenti o molto carenti le informazioni sull'edizione cartacea di riferimento e sulle operazioni compiute per produrre la versione digitale.

Una base di dati testuale dei testi latini tardi che sia completa, affidabile, aggiornata dal punto di vista delle tecnologie informatiche, e aperta, non esisteva. Uno studioso scrisse nel 2010 in un *referee* a proposito del progetto, «no reasonably usable database of these texts presently exists»⁴.

raccolta esiste, allo stesso modo che in una collana di testi a stampa i singoli volumi sono fisicamente distinti, ma condividono tutti una medesima impostazione tipografica, medesimi criteri scientifici di massima per quanto riguarda introduzioni, traduzioni se presenti, apparato.

³ Il numero 1 si apriva con un'intervista a Kraft che già nel 1984 presentava l'annosa questione degli strumenti informatici che spoglierebbero gli studiosi delle loro prerogative: «[Q: Doesn't the introduction of computers into humanistic scholarship endanger the intellectual and poetic freedom that is typical of that perspective?] A: We are talking about a tool for research and writing, something that takes some of the drudgery out of parts of the humanist's work. The computer [...] gives you new perspectives on the material with which you are working[...].», <http://ccat.sas.upenn.edu/gopher/other/journals/offline/Offline%20vols.%201-36>.

⁴ Dei testi latini tardi, dal II al V secolo d.C. non esiste nemmeno un elenco completo e consolidato. Una delle attività iniziali del gruppo di ricerca che guida la biblioteca digitale è proprio la definizione dettagliata di tale elenco, operazione che peraltro, come la storia del

digilibLT Biblioteca digitale di testi latini tardoantichi

Home | Il progetto | Notizie | Tardoantico nel web | Aiuto | Contatti e feedback | Accedi

Ricerca

Opere disponibili per la ricerca
Ricerca avanzata
Ricerca bibliografica

Letture e download di testi

Opere
Autori
Bibliografia

Per ordine cronologico
II III IV V VI VII
Datazione incerta
Elenco autori

Per ordine alfabetico
A B C D E F G H I L M N O
P Q R S T V
Elenco autori

Risorse

Opere sul tardoantico
Canone autori tardoantichi
Download di font e software
Scarica i testi in formato PHI /TLG

Proponente aggiuntivo

Biblioteca digitale di testi latini tardoantichi

Il progetto digilibLT fornisce un canone degli autori e delle opere tardoantichi e un database di testi, completo e chiaro nell'indicazione delle edizioni adottate come riferimento e degli interventi editoriali eventualmente operati su di essi, di libera e gratuita consultazione online. Le maschere di consultazione consentono ricerche testuali sull'insieme dei testi, su loro raggruppamenti (per autore, per epoca, per tipologia testuale), oltre che per singoli autori e opere. I testi sono anche resi disponibili al download, in modo che ogni studioso possa scaricare le opere di suo interesse e lavorare su di esse per le proprie ricerche con la massima flessibilità. I testi sono codificati sia secondo lo standard TEI, sia secondo lo standard Beta coding, così da permetterne l'uso con i programmi diffusi per l'interrogazione dei cdrom del TLG e del PHI. Il sito si completa con schede illustrative di autori e opere, bibliografie degli studi e schede catalografiche, opere su autori tardoantichi in pdf. [R.Tabacco]

Un nuovo testo retorico
25/4/2012
Su DigilibLT sono stati pubblicati i *Præcepta artis rhetoricae* di Giulio Severiano, un testo particolarmente significativo per la retorica tardoantica. [Andrea Balbo]

Due nuovi testi disponibili
27/2/2012
Sono disponibili sul sito il frammento del dialogo *Vergilius orator an poeta* attribuito a P. Anno Floro (II secolo) e l'*Opusculum* di Giulio Esuperanzio (IV-V secolo). [Andrea Balbo]

Nuove accessioni su DigilibLT
1/2/2012
Sono disponibili nel sito DigilibLT la *Gratiarum actio* di Ausonio (IV secolo), due testi medici, le *Celeres passiones* e le *Tardae passiones* di Celio Aureliano (V secolo) e la *Philosophiae Consolatio* di Boezio (VI secolo). [Andrea Balbo]

Figura 2. Home page di digilibLT

Realizzare l'idea richiedeva mezzi economici ingenti, che non si vedeva come raccogliere. Poi nell'anno 2008 la Regione Piemonte pubblicò un *Bando regionale per progetti di ricerca in materia di scienze umane e sociali per finanziare ricerche da effettuarsi negli atenei piemontesi nell'ambito delle scienze umane e sociali*. Con alcuni colleghi del Dipartimento (L. Battezzato, R. Piastrì, G. Vanotti⁵) e con la direttrice della biblioteca del Dipartimento (Silvia Botto) si decise di concorrere presentando l'idea attraverso un progetto redatto da chi scrive. Il progetto per la creazione della biblioteca digitale del latino tardo denominata digilibLT (Digital Library of Late-Antique Latin Texts) venne approvato dalla Regione Piemonte nel-

TLG insegna, non si conclude tanto rapidamente. Come orientamento di lavoro si partirà dai testi degli scrittori pagani, prosa e poi poesia, per passare in seguito agli scrittori cristiani. Questi ultimi sono peraltro già disponibili in una raccolta completa (la *Patrologia Latina* del Migne pubblicata da Chadwyck-Healey) che, benché proprietaria, chiusa, a pagamento e per molti testi e autori decisamente superata come qualità delle edizioni, è pur sempre una raccolta completa.

⁵ Al gruppo vercellese dell'Università del Piemonte Orientale si unirono nel progetto alcuni colleghi del Dipartimento di Studi Umanistici dell'Università di Torino: A. Balbo, V. Lomanto, E. Malaspina; ad essi si aggiunsero poi in varie forme, A. Borgna, F. Ciotti, L. Mosca, S. Musso, N. Rosso.

l'agosto 2009, risultando il secondo nella graduatoria finale che ne comprendeva più di settanta⁶. La biblioteca è ora attiva e operante sotto la direzione di Raffaella Tabacco e di chi scrive, accessibile all'indirizzo <<http://digiliblt.unipmn.it>> (fig. 2). I testi latini di cui si parla sono noti agli studiosi attraverso edizioni di editori d'oggi, in generale edizioni critiche cioè edizioni del testo accompagnato da note che descrivono il modo in cui esso ci è testimoniato dai principali codici manoscritti (qualche volta anche papiri).

Nelle edizioni critiche recenti il testo dell'opera è libero da diritti perché l'autore antico è morto da più di settant'anni, mentre sono protetti tutti gli apparati di commento (introduzione, apparato critico, commento vero e proprio) in quanto opera di studiosi che in genere non sono morti da più di settant'anni (perché la produzione di edizioni critiche delle opere classiche è continua). La biblioteca digitale digilibLT si è quindi orientata con chiarezza fin dall'inizio sul raccogliere nelle sue edizioni digitali il "testo stabilito", come viene chiamato tecnicamente, e sul lasciare da parte tutti gli apparati di commento protetti da diritti. Parrebbe ovvio che debba essere così, ma non sempre accade. Un'altra scelta di principio venne fatta fin dalla formulazione del progetto presentato alla Regione Piemonte: che la biblioteca adottasse i principi dell'accesso aperto e che fosse posta tutta intera sotto una precisa ed esplicita licenza d'uso.

Questioni aperte e scelte operate

Formulare il progetto digilibLT e costruire la biblioteca digitale ha significato prendere posizione su un complesso di questioni metodologiche e di principio che sono interessanti e significative anche al di fuori dello studio del latino e che vengono qui esposte con l'intento di descrivere in modo compiuto e organico una realizzazione pratica che si ritiene possa risultare alla prova dei fatti anche una pratica valida.

Tre sono le specificità di digilibLT che occorre tenere presenti:

- la prima è che si tratta di una biblioteca digitale di testi di opere letterarie (e non, per esempio, di articoli di rivista);
- la seconda è che le opere in questione sono libere da diritti;
- la terza è che i testi delle opere sono digitalizzati ad hoc.

Altre biblioteche digitali possono raccogliere articoli di rivista già digitalizzati e protetti da diritti, oppure opere che vengono digitalizzate parzialmente (vengono prodotte le immagini delle pagine, ma non viene operato il riconoscimento dei caratteri) e in tali casi le questioni progettuali e di sistema che occorre affrontare possono variare anche di molto.

⁶ http://www.regione.piemonte.it/innovazione/images/stories/ricerca/dwd/sc_umane_all1.pdf/.

Accesso aperto, licenze Creative Commons, software libero

Accesso aperto

L'accesso aperto, cioè libero, alle risorse digitali ha una valenza e un significato particolare nel caso delle risorse prodotte dalla ricerca scientifica: dagli articoli pubblicati su riviste alle monografie, ai prodotti della ricerca in generale. Se «i mercati sono conversazioni tra esseri umani»⁷, la ricerca non lo è di meno e infatti essa ha due ambiti di discorso: quello in cui i ricercatori parlano tra loro, e quello in cui parlano con la società. La comunicazione libera, ampia e aperta tra i ricercatori non è particolarmente favorita dall'attuale e dominante sistema delle pubblicazioni, per cui si può arrivare fino al punto che lo studioso non può dare a un collega una copia di un suo articolo pubblicato su una rivista scientifica: cosa corretta dal punto di vista della tipologia dominante dei contratti di edizione, ma abnorme dal punto di vista della circolazione e dello sviluppo del sapere. Ma non si deve dimenticare l'altro discorso, quello che si svolge tra ricercatori e società: l'orizzonte in cui si colloca la ricerca è un orizzonte sociale, la ricerca ha bisogno di essere conosciuta, accolta, accettata e "utilizzata" dalla società, anche solo perché è la società che in vari modi e attraverso vari soggetti finanzia la ricerca. A maggior ragione se si tratta di ricerca finanziata con fondi pubblici: come un *venture capital* di cui sono soci tutti i cittadini che pagano le tasse e che in quanto finanziatori hanno diritto di conoscere a quali conclusioni sia giunta la ricerca che hanno finanziato.

L'iniziativa per l'accesso aperto ha pubblicato il suo primo documento nel 2002, *Budapest Open Access Initiative*⁸, per impulso della Open Society Foundation di George Soros, cui hanno fatto seguito nel 2003 la *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*⁹, e infine – per quanto concerne l'Italia – nel 2004 la *Dichiarazione di Messina*¹⁰. Poiché il centro dell'attenzione è sugli articoli pubblicati dalle riviste scientifiche, le pratiche più correntemente connesse con l'accesso aperto sono l'autoarchiviazione in spazi Web personali o istituzionali dei pre-print o degli articoli e la pubblicazione degli articoli stessi in riviste scientifiche ad accesso aperto. Quindi la costruzione di una biblioteca digitale, tanto più se di testi letterari non contemporanei, sembrerebbe solo tangenzialmente toccata dalle problematiche in questione. Ma in realtà, appunto, al di là delle pratiche più diffuse e che attirano maggiormente l'attenzione, qualsiasi progetto di ricerca che arrivi a un prodotto finale è suscettibile di accesso aperto o comunque deve prendere posizione sulla questione se rendere liberamente accessibili oppure no i suoi risultati e i suoi prodotti.

⁷ *The Cluetrain Manifesto*, <http://cluetrain.com/#manifesto>.

⁸ <http://www.soros.org/openaccess/read>.

⁹ <http://oa.mpg.de/berlin-prozess/berliner-erklarung/>.

¹⁰ Propriamente *Documento italiano a sostegno della Dichiarazione di Berlino sull'accesso aperto alla letteratura accademica*, <http://www.aepic.it/conf/Messina041/index981f.html>.

Licenze Creative Commons

Connessa con il tema dell'accesso aperto è la questione delle condizioni d'uso dei prodotti resi liberamente accessibili. Accesso aperto non significa che il lettore fruitore possa fare ciò che vuole della pubblicazione con cui è entrato in contatto. Nel mondo della carta i diritti di proprietà intellettuale sulle opere dell'ingegno sono dell'autore dell'opera stessa che ne concede poi lo sfruttamento economico a un editore (nel caso delle opere a stampa), sfruttamento esclusivo, in genere. Il vantaggio che ne viene all'autore è che la sua opera circola attraverso le copie del libro o della rivista che l'editore provvede a distribuire e che dalla circolazione e vendita deriva all'autore un vantaggio economico. Ma se, come sta accadendo da un po' di tempo a questa parte, per chi pubblica risultati di ricerca la circolazione si riduce e i proventi economici non esistono, allora è chiaro che il modello precedente non funziona più.

Si possono ricordare due vicende, una italiana e una statunitense. Quella statunitense riguarda la Harvard University: il Faculty Advisory Council, il 24 aprile 2012 scrive:

«Many large journal publishers have made the scholarly communication environment fiscally unsustainable and academically restrictive. This situation is exacerbated by efforts of certain publishers (called "providers") to acquire, bundle, and increase the pricing on journals. ... Prices for online content from two providers have increased by about 145% over the past six years, which far exceeds not only the consumer price index, but also the higher education and the library price indices. ... major periodical subscriptions, especially to electronic journals published by historically key providers, cannot be sustained: continuing these subscriptions on their current footing is financially untenable¹¹.»

Quella italiana riguarda, a titolo di esempio di una tendenza generale del mercato, il caso di una rivista scientifica di area umanistica il cui prezzo dal 1998 a oggi è cresciuto del 651%, come ha dettagliatamente esposto Claudio Giunta prima su «La Rivista dei Libri» e poi su «Il Sole 24 Ore»¹².

Senza dimenticare il matematico statunitense Tim Gowers e il suo invito ampiamente argomentato¹³ rivolto ai colleghi a sottoscrivere una dichiarazione di boicottaggio di Elsevier¹⁴ in quanto l'editore carica di costi esageratamente alti le sue riviste e pratica politiche commerciali vessatorie, invito che aveva raccolto, alla data di maggio 2012, 11.500 adesioni. L'azione ha avuto come primo risultato che l'editore ha risposto pubblicamente alla comunità dei matematici¹⁵ e ha parzial-

¹¹ <http://isites.harvard.edu/icb/icb.do?keyword=k77982&tabgroupid=icb.tabgroup143448>.

¹² Claudio Giunta, *Quanto (ci) costa l'editoria accademica?*, «la Rivista dei libri», 20 (2010), n.2, p. 30-34, <http://www.larivistadeilibri.it/2010/02/giunta.html>; Claudio Giunta, *Le riviste per soli ...ricchi*, «Il Sole 24 Ore», 15 aprile 2012, p. 31.

¹³ <https://gowers.wordpress.com/2012/01/21/elsevier-my-part-in-its-downfall/>.

¹⁴ <http://www.thecostofknowledge.com>.

¹⁵ http://www.elsevier.com/wps/find/P11.cws_home/lettertothecomunity.

mente cambiato alcune politiche¹⁶.

Tornando al tema principale: se il modello tradizionale non funziona più bene come prima e gli autori desiderano cambiare il modo e la forma di pubblicazione delle loro ricerche, non perciò rinunciano a tutto, a tutti i vincoli e a tutti i diritti: si rinuncia ad «alcuni» vincoli e diritti sulle proprie opere, non più *all rights reserved*, ma *some rights reserved*. Si collocano in questo orizzonte le licenze Creative Commons (il nome è correntemente abbreviato in CC), frutto del lavoro del giurista di Harvard Lawrence Lessig. Un autore che decida di porre sotto licenza CC il suo lavoro ha molte possibili licenze tra cui scegliere, ma tutte condividono il vincolo dell'*attribuzione* cioè un autore non può rinunciare a essere riconosciuto come autore di ciò che ha prodotto:

Attribuzione

Attribuzione - Non opere derivate

Attribuzione - Non commerciale - Non opere derivate

Attribuzione - Non commerciale

Attribuzione - Non commerciale - Condividi allo stesso modo

Attribuzione - Condividi allo stesso modo

Il loro senso è chiaro già dalla formulazione sintetica che le descrive, ma per una descrizione dettagliata, completa di altre licenze meno usuali, si può visitare il sito web¹⁷. A parte l'attribuzione, in sostanza l'autore sceglie se permettere o non permettere l'utilizzo a fini commerciali del proprio prodotto e se vuole impedire la produzione di altre opere a partire dalla sua; e se vuole che la licenza con cui distribuisce il prodotto vincoli anche i successivi utilizzatori. In altri termini le licenze CC implicano che i prodotti tutelati possano essere copiati e ridistribuiti ma definiscono e pongono dei vincoli per questa circolazione. Particolarmente adatta a una biblioteca digitale è la licenza Attribuzione - Non commerciale - Condividi allo stesso modo, che garantisce il riconoscimento dell'autore (Attribuzione), vincola il fruitore a un uso senza fini di lucro (Non commerciale; ma l'uso commerciale può essere contrattato separatamente) e vincola chi ridistribuisce l'opera a ridistribuirla con gli stessi vincoli (Condividi allo stesso modo). Essa prevede la libertà per chi scarica i testi (ma in senso lato per chiunque utilizzi il sito del progetto) «di riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare» nonché «di modificare l'opera», ad alcune condizioni: l'utilizzatore

¹⁶ Ha per esempio ritirato il sostegno al Research Works Act http://www.elsevier.com/wps/find/intro.cws_home/newmessengerwa, che nel quadro delle politiche protezionistiche statunitensi SOPA e PIPA intende vietare la pubblicazione ad accesso aperto di ricerche finanziate con fondi federali.

¹⁷ <http://www.creativecommons.it/Licenze> e anche «DigItalia» 4 (2009), n. 1: Deborah De Angelis, *Brevi note in tema di applicabilità delle licenze creative commons [...]*, p. 9-23.

«deve attribuire la paternità dell'opera nei modi indicati dall'autore o da chi gli ha dato l'opera in licenza e in modo tale da non suggerire che essi avallino lui o il modo in cui egli usa l'opera»; «non può usare l'opera per fini commerciali»; «se altera o trasforma l'opera, o se la usa per crearne un'altra, può distribuire l'opera risultante solo con una licenza identica o equivalente a quella originaria».

Una biblioteca digitale come digilibLT, che distribuisce testi liberi da diritti, pone sotto licenza CC le opere:



e quindi se uno studioso scarica dalla biblioteca digitale digilibLT alcuni testi per studiarli con un programma di sua scelta, occorre che egli citi tra le fonti della sua ricerca il progetto digilibLT e le persone che hanno curato il o i testi che egli ha utilizzato, esattamente come accadrebbe per i libri; non è permesso invece che qualcuno scarichi i testi e li utilizzi tutti o in parte per una pubblicazione che venga venduta e gli procuri un lucro. Ma digilibLT pone sotto licenza CC anche le pagine Web del sito:



basta pensare al fatto che un sito nel suo complesso e le pagine che lo compongono una per una sono frutto anch'essi di lavoro intellettuale; chi volesse potrebbe creare un sito simile a digilibLT prendendo spunto dai suoi contenuti, ma sarebbe tenuto a dichiarare per ogni pagina Web che all'origine del suo sito c'è il sito digilibLT. digilibLT richiede che chi desideri scaricare uno o più testi si identifichi (attraverso una registrazione che chiede alcuni dati essenziali: nome e cognome, eventuale affiliazione, indirizzo di posta elettronica, nonché l'approvazione esplicita della licenza Creative Commons sotto la quale sono posti il sito e i contenuti). Fatto questo, la persona sarà libera di scaricare, riutilizzare e anche modificare i testi scaricati, alle condizioni definite dalla licenza.

Chi sceglie di distribuire contenuti digitali sotto licenza CC ovviamente e contemporaneamente decide di non adottare “protezioni” per i suoi prodotti digitali: nessun sistema di criptazione e frammentazione dei dati¹⁸. Anche a ritenere che ci siano ragioni per volerli adottare, i sistemi di criptazione e di frammentazione sono inevitabilmente legati alla tecnologia hardware/software del tempo in cui nascono; ma l’evoluzione della tecnologia è così rapida che nel giro di una decina d’anni al massimo nemmeno il legittimo utente è più in grado di accedere ai suoi dati: si tratta di un’esperienza che molti hanno fatto e fanno talora con sorpresa. Aver pagato tutto il dovuto e comportarsi bene a norma di legge e regolamenti non garantisce di poter fruire nel tempo del prodotto digitale acquistato. In tale situazione c’è un danno per chi ha acquistato l’opera (o il diritto di fruirne, come spesso accade per i prodotti digitali) ma anche per chi l’ha prodotta, in quanto non può essere conosciuto per mezzo di essa; per non parlare del fatto che se l’opera – ove disponibile – fosse stata acquistata in formato analogico, su supporto fisico, la fruizione avverrebbe senza vincoli di sorta, senza limitazioni per decine e decine di anni (secoli, nel caso di libri o riviste). Le licenze CC offrono una fruizione durevole nel tempo, in quanto non adottano artifici informatici, dando così un piccolo aiuto all’utente per superare i problemi causati dall’evoluzione della tecnologia, per esempio nel formato dei file.

Software libero

Si presenta a questo punto il tema del software libero (*free software*, che non è da confondere con *open source software* anche se nel parlare corrente le due espressioni sono spesso usate come equivalenti) che ha come corollario il tema dei formati aperti dei dati. Il software libero non presenta costi d’uso per la licenza ed è aperto alla possibilità di modifiche per renderlo più funzionale ai propri scopi. In digilibLT tutto il software è *free*: il sistema operativo, il *content management system*; il motore di ricerca testuale; il gestore del *repository*; le *utility*. Anche se permette la lettura di dati in formati proprietari, il software libero coerentemente con il principio guida ne prevede il salvataggio in formati aperti e liberi, spesso basati su XML nel caso dei testi. Il senso della scelta è duplice: i dati si possono leggere anche in assenza del programma che li ha prodotti; i dati sopravvivono meglio nel tempo (sono resilienti) perché XML è un formato basato su testo e quindi è leggibile da qualsiasi editor. Se i dati testuali di un’opera distribuiti sotto licenza CC fossero scritti in un formato proprietario (si pensi a titolo di esempio all’onnidiffuso formato DOC di Word), si porrebbe un problema di libertà e trasparenza di accesso nonostante la licenza CC perché se il programma che li legge tra vent’ an-

¹⁸ La frammentazione è adottata tipicamente da banche dati testuali su CD-ROM o DVD: i dati si possono copiare e costituiscono un unico grande file dentro il quale sono divisi in piccoli blocchi tutti disordinatamente mescolati fra loro senza che nulla li identifichi visibilmente; ed è solo il programma di consultazione fornito insieme ai dati che permette di gestirli correttamente.

ni non esistesse più¹⁹, i dati scritti nel suo formato non sarebbero più leggibili, pur essendo posti sotto una licenza che ne permette e ne favorisce la circolazione.

Preparazione dell'edizione digitale

Canone

Attualmente non esiste un elenco completo e preciso che per il periodo menzionato citi tutti gli autori latini pagani ad oggi noti insieme con le loro opere, perché il repertorio più autorevole in materia²⁰, cioè la lista delle opere utilizzate per la creazione del *Thesaurus Linguae Latinae*, non è recentissimo e opere di non grande mole, frutto di acquisizioni recenti, sono state talora pubblicate in riviste scientifiche anziché in volumi a stampa. Pertanto il lavoro scientifico inizia con la redazione di tale elenco. Nel sito esso è disponibile attraverso un'apposita voce di menu denominata *Canone degli autori latini* in cui per ogni autore e opera è indicata l'edizione critica di riferimento adottata dalla biblioteca. Il termine canone potrebbe dare a intendere una volontà in qualche modo normativa, ma in realtà si tratta solo di una ripresa per analogia dal TLG Canon, il canone del TLG, cioè l'elenco pubblicato nel 1990²¹ delle opere contenute nel CD-ROM dei testi greci classici²² curato dall'omonimo laboratorio dell'Università di California.

Digitalizzazione del testo

Se si vuole rendere disponibile nel mondo digitale un'opera stampata su carta occorre intraprendere molteplici operazioni. La più semplice consiste nel produrre un'immagine digitale delle pagine a stampa²³, ma è chiaro che in tal modo non si fa molta strada. Un passo successivo più attento alla fruizione consiste nell'impacchettare le immagini nel giusto ordine all'interno di un file PDF che funge da contenitore. Nell'operazione di impacchettamento il programma che gestisce il formato PDF²⁵ può operare il riconoscimento ottico dei caratteri (OCR). Nel primo caso (semplice impacchettamento delle immagini dentro un file PDF) non si possono effettuare ricerche sul testo, si può solo leggerlo; nel secondo caso si possono fare ricerche sul testo pur continuando a leggerlo nelle immagini delle pagine. In entrambi i casi non si tratta però di vere edizioni digitali dei testi.

¹⁹ Tecnicamente potrebbe esistere un altro programma capace di leggere il formato in questione.

Peraltro i formati proprietari come quello in esame sono pubblicamente noti "di fatto", ma non ne esiste documentazione pubblica ufficialmente rilasciata; e in ogni caso il proprietario non si vincola alle specifiche che ha pubblicato.

²⁰ *Thesaurus linguae Latinae. Index librorum scriptorum inscriptionum ex quibus exempla afferuntur*, a cura di Dietfried Krömer e Cornelis G. van Leijenhof, Leipzig: Teubner, 1990.

²¹ *Thesaurus Linguae Graecae Canon of Greek Authors and Works*, a cura di Luci Berkowitz e Karl A. Squitier, New York - Oxford: Oxford University Press, 1990.

²² Il TLG CD-ROM A iniziò a essere distribuito dal 1985, il CD-ROM E dal 2000.

²³ La riproduzione per immagini di un'opera a stampa è libera se essa non è più coperta da diritti di sfruttamento economico della proprietà intellettuale, che decadono dopo settant'anni dalla morte dell'autore.

²⁴ In genere si pensa ad Acrobat ma esistono anche altri programmi che svolgono le stesse funzioni.

L'acquisizione delle immagini di digilibLT avviene per mezzo di uno scanner (BookdrivePro di Atiz) capace di lavorare fino al formato doppio A2, che utilizza un leggio a V e apparecchi fotografici in luogo del normale letto piano. La gestione delle pagine, molto rapida e priva di difficoltà, avviene con Bookdrive Capture, un software fornito dal produttore dello scanner, che permette di definire o modificare via cavo le impostazioni degli apparecchi fotografici necessarie alla migliore resa nella cattura delle immagini (esposizione, diaframma, sensibilità ISO, dimensione e tipo dell'immagine, ...). Successivamente per mezzo di Bookdrive Editor, anch'esso fornito dal produttore dello scanner, le immagini delle pagine vengono se necessario raddrizzate, scontornate, ripulite da eventuali dominanti di colore indesiderate (l'ingiallimento della carta, per esempio) e impacchettate a scelta dell'utente o in un file TIFF multipagina, o in un file PDF che può contenere a scelta immagini TIFF o JPEG. Il file TIFF multipagina, o PDF, viene poi passato a un programma di riconoscimento ottico dei caratteri in cui si seleziona la parte della pagina sui cui lavorare, cioè il testo stabilito, per la generazione di un file di testo puro. Il file PDF con le immagini a questo punto viene cancellato.

Correzione del testo

La digitalizzazione del testo a partire da pagine aperte su un leggio a V anziché da pagine che si incurvano nella costola quando sono squadernate su un piano dà luogo a discreti vantaggi nella digitalizzazione del testo. Gli errori di riconoscimento non vengono però eliminati del tutto, e quindi occorre che il testo digitalizzato attraversi una o due fasi di correzione. La correzione è molto tradizionale: si confronta il testo digitale con il testo di partenza.

Si potrebbe anche immaginare una situazione con due differenti file di testo generati da due differenti programmi di OCR; successivamente i due file di testo vengono confrontati fra loro con un programma di tipo *diff* allo scopo di evidenziare le discrepanze tra i due file, che con grande probabilità corrispondono a errori di digitalizzazione; il file così ottenuto fondendo i due sorgenti ed evidenziando le discrepanze viene letto da una persona che conosce il latino e che può quindi correggere gli errori. L'ipotetica modalità di correzione del testo così sinteticamente descritta è di fatto una rivisitazione del *double-keying*, cioè la duplice battitura alla tastiera per ogni testo, cui segue il confronto fra i due testi per evidenziare le discrepanze che con grande probabilità saranno punti in cui uno dei due testi contiene un errore. La battitura del testo all'estero (in India, o Corea, o Cina, di solito) anche con garanzia di un risultato privo di errori, avrebbe eliminato la spesa per lo scanner e per il software di OCR, ma da un lato avrebbe avuto comunque un suo costo specifico (per giunta inestricabilmente connesso con la preoccupazione di trarre vantaggio dallo sfruttamento del lavoro di altre persone) e dall'altro non avrebbe eliminato la necessità, per il gruppo di ricerca, di leggere ogni singola pagina di testo per accertarsi che essa fosse effettivamente priva di errori. La scelta di lavorare in proprio con uno scanner molto evoluto ha avuto origine anche da

una specifica scelta di gestione: se l'istituzione riceve un finanziamento corposo è preferibile fare in modo che ciò lasci beni e attrezzature piuttosto che spendere nell'acquisto di servizi.

Marcatore del testo (annotazione)

Produrre l'edizione digitale di un'opera a stampa significa produrre un file che contenga il testo dell'opera (cioè byte-carattere, non byte-pixel) accompagnato da informazioni codificate. Infatti il testo che viene digitalizzato, in quanto viene separato dal supporto cartaceo al quale la nostra storia culturale lo ha stabilmente e fortemente associato, se non riceve qualche tipo di specifica cura editoriale rimane indebolito e impoverito. Il puro e semplice fatto di produrre un file di testo a seguito di operazioni di riconoscimento ottico dei caratteri non dà luogo a un'edizione digitale dell'opera. Fin dagli albori della digitalizzazione di testi classici, imprese quali il CD-ROM TLG e il CD-ROM del PHI adottarono una loro codifica del testo digitale, nota come *Beta code*²⁵, che permetteva tra l'altro ai programmi di analisi testuale dedicati (quasi in ordine cronologico di apparizione LBase, TLG Workplace e PHI Workplace, SNS Greek, Musaios, Lexis, Lector, Diogenes) di riconoscere la strutturazione formale del testo in unità di vario livello gerarchico (libro, capitolo, paragrafo, per intendersi; ma con tutte le variazioni effettivamente esistenti nella tradizione testuale di ogni opera) così che chi cercava una parola otteneva dal programma non solo i passi in cui la parola si trova, ma anche gli estremi del passo.

Quell'esigenza non è scomparsa, ma anzi si è specificata e dettagliata e per darvi risposta prese avvio nel 1988 un'iniziativa dal basso denominata TEI, Text Encoding Initiative²⁶, che in un documento periodicamente aggiornato noto come TEI Guidelines²⁷, indica come descrivere il testo e l'opera²⁸ in termini formalizzati (e anche comprensibili alle macchine), la struttura dei testi utilizzando il linguaggio di marcatore noto come XML²⁹. Per mezzo dei marcatori³⁰ in formato TEI è possibi-

²⁵ <http://www.tlg.uci.edu/encoding/BCM2010.pdf> per il *TLG® Beta Code Manual 2010* e <http://www.tlg.uci.edu/encoding/quickbeta.pdf> per il *TLG Beta Code Quick Reference Guide*.

²⁶ The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics», <http://www.tei-c.org/>.

²⁷ <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>; la prima versione è del 1990.

²⁸ Pur senza troppo dettaglio, occorre ricordare che esistono due tipologie di *markup*: procedurale e descrittivo. Il *markup* procedurale è orientato alla produzione a stampa del testo, ne sono esempi i linguaggi TeX e LaTeX. TEI è invece un *markup* orientato alla descrizione del testo (il che non impedisce che con opportune trasformazioni un *markup* descrittivo TEI-XML possa dare luogo alla produzione di una versione a stampa del testo).

²⁹ «Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere», <http://www.w3.org/XML/>.

³⁰ *Markup*, *encoding*, *codifica*, *annotazione*, *marcatore*, e altri ancora, sono i termini - parzialmente

le descrivere l'opera nella sua versione a stampa di riferimento, indicare in quale lingua è scritta, quale set di caratteri adottati, quale sia il suo livello di correttezza e revisione, chi vi abbia lavorato, se e dove vi siano dei punti del testo rovinati o corretti dagli studiosi, se e dove il testo contenga delle immagini ... Questa lista potrebbe estendersi per pagine e pagine. È talmente ampio e dettagliato l'insieme delle informazioni codificate dal manuale TEI, giunto da poco alla sua versione 5³¹, che ne è stato ricavato un sottoinsieme noto come *TEI Lite*³² costituente per così dire l'ossatura portante, l'insieme minimo. Poi in realtà, come accade spesso, ogni progetto e ogni tipologia di testi ha bisogno, oltre alle informazioni (oltre ai *tag*) che costituiscono il sottoinsieme minimo *TEI Lite*, di informazioni (*tag*) che fanno parte dello standard completo³³. La codifica TEI adottata da digilibLT, che per uso interno è stata descritta in un conciso manuale, avviene utilizzando un ambiente di editing XML specificamente orientato a TEI, quale è <Oxygen/>. Dalla versione del testo con codifica TEI vengono ricavate tutte le altre versioni/formati disponibili nel sito.

Beta code

Beta code è una codifica di carattere e di formattazione sviluppata specificamente per permettere una rappresentazione digitale accurata dei testi greci classici (e anche di altre lingue antiche) su un sistema basato su caratteri ASCII. Fu sviluppato da D. Packard alla fine degli anni 70 e adottato dal TLG nel 1981. Beta è diventato lo standard per codificare il greco polifonico ed è stato usato anche in numerosi altri progetti come Perseus, [le banche dati testuali del] Packard Humanities Institute, i Duke documentary Papyri, il Greek Epigraphy Project³⁴.

sinonimi – che indicano l'insieme delle operazioni con cui in un testo vengono inserite in modo formalizzato informazioni sul testo stesso. In particolare si nota una tendenza all'uso del termine annotazione laddove le informazioni riguardanti il testo sono centrate sulla semantica.

³¹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

³² *TEI Lite* is a specific customization of the TEI tagset, designed to meet "90% of the needs of 90% of the TEI user community". Due to its simplicity and the fact that it can be learned with relative ease, TEI Lite has been widely adopted, particularly by beginners and by big institutional projects that rely on large teams of encoders to markup their documents», <http://www.tei-c.org/Guidelines/Customization/Lite/>. La sua versione più recente è il manuale *TEI Lite: Encoding for Interchange: an introduction to the TEI — Revised for TEI P5 release* di Lou Burnard, C. M. Sperberg-McQueen (Febbraio 2006), <http://www.tei-c.org/release/doc/tei-p5-exemplars/html/teilight.doc.html>. In italiano è disponibile *on-line* la versione del 1998 *TEI Lite: introduzione alla codifica dei testi*, traduzione italiana Fabio Ciotti, Guendalina Demontis, Giuseppe Gigliozzi, Massimo Guerrieri, Andrea Loreti, revisione e cura traduzione italiana di Fabio Ciotti, <ftp://ftp-tei.uic.edu/pub/tei/lite/teiu5-it.html>; a stampa è disponibile il volume *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari*, a cura di Fabio Ciotti, Milano: Sylvestre Bonnard, 2005.

³³ Il dibattito, nel mondo degli studi che si interessa di questi temi, è vivace e mostra un numero crescente di studiosi che sostengono la necessità di un approccio "laico" e non "osservante" all'impiego della codifica TEI.

³⁴ Dalla pagina iniziale del *TLG® Beta Code Manual 2010* cit.

Un esempio aiuta a capire di che cosa si parla: in un testo greco codificato in Beta code, la sequenza di segni ASCII:

*S3

genera in un programma capace di interpretarla la visualizzazione del carattere noto come "sigma lunato maiuscolo" che si presenta con questa forma:

Σ

In Beta code, una riga che contiene informazioni sulla struttura del testo è identificata dal fatto che il primo carattere è una tilde (~). I vari livelli di struttura testuale – chiamati campi – sono rappresentati da lettere minuscole, e una lettera può essere seguita dal valore (numerico, o stringa ASCII) da assegnare al campo. I livelli a e b sono numerici, riservati rispettivamente all'autore e al titolo dell'opera, così come indicati nel canone della raccolta testuale, mentre il livello c contiene il nome abbreviato dell'opera. Quindi una riga come questa:

~a"0001"b"001"c"Arg"

con a"0001" rimanda all'autore che nel canone corrisponde al n. 1, Apollonio Rodio; con b"001" rimanda all'opera numero 1 tra quelle di Apollonio Rodio, cioè le Argonautiche; e c"Arg" indica l'abbreviazione del titolo dell'opera. Analogamente avviene per i livelli inferiori della struttura del testo (libri, versi, capitoli, pagine eccetera).

TEI

Nonostante la sua complessità, qui appena accennata, il Beta code non è pensato per, e quindi non è in grado di, rappresentare nel testo digitale la complessità del testo su carta. Per esempio non prevede la possibilità di inserire nel file stesso indicazioni sull'edizione di riferimento, sul curatore della versione digitale nella quale anzi TEI riconosce una vera e propria edizione digitale. Né di identificare elementi strutturali, sintattici, linguistici, quali discorso diretto, citazione indiretta, forme verbali anomale, eccetera, tutte marcature possibili con TEI. La codifica TEI però è complessa, più complessa del Beta code, ma soprattutto non esistono oggi programmi paragonabili a Musaios o Diogenes che permettano a uno studioso di interrogare con vantaggio una base di dati testuali sfruttando specificamente il fatto che essa sia stata marcata secondo le linee guida TEI; fa eccezione CATMA³⁵, uno strumento per lavorare sui testi sia codificandoli secondo le linee guida TEI sia utilizzando una codifica TEI già presente nel testo, che è distribuito dall'Università di Amburgo.

Ci si può chiedere perché fare tutto questo, cioè perché adottare (coloro nel mondo che lo fanno, e nello specifico la biblioteca digilibLT) una codifica complessa e anche ostica per chi non ne sia un vero specialista, difficile da gestire senza il ricorso ad un ambiente software molto specialistico (<oxygen/>), non diffusa tra

³⁵ <http://www.slm.uni-hamburg.de/catma/index.html>.

gli studiosi quanto lo è attualmente il Beta code per mezzo dei CD-ROM TLG e PHI. Al di là del fatto che se in quella direzione muove, comunque, il mondo della basi di dati testuali a meno di forti e fondate obiezioni non conviene “tirarsi fuori”, la ragione più interessante è la grande flessibilità per cui seguendo le linee guida TEI non solo è possibile codificare/rappresentare qualsiasi struttura testuale, ma è anche possibile codificare elementi di contenuto che risultino interessanti per lo studioso: le domande in discorso diretto, le apostrofi al lettore, il tema luce-oscurità, tutto ciò che si vuole insomma. Così la “marcatura” svela il suo carattere di “annotazione”: grazie ad essa lo studioso di fatto riversa nel testo, in maniera formalizzata e riutilizzabile da lui stesso o da altri, la sua conoscenza del testo stesso. Non diversamente da quanto avviene quando all’interno di un materiale testuale vengono costruiti dei collegamenti ipertestuali, che svelano la rete delle connessioni di significato presenti in tale materiale secondo il punto di vista di chi ha creato i link. Ma la codifica del contenuto secondo chiavi di lettura di interesse dello studioso ha un elemento in più di interesse: è molto chiara e quindi riusabile; mentre i link ipertestuali così come il web ce li ha fatti conoscere non sono tipizzati³⁶ e dunque non permettono di esplicitare pienamente il senso del collegamento che viene in parte lasciato da ipotizzare e scoprire al lettore/navigatore. Si può quindi pensare la codifica del contenuto da parte dello studioso come un lavoro di glossatura del testo che non avviene su carta con la penna, bensì si colloca in forma codificata all’interno del file che contiene il testo digitalizzato.

Perché la codifica in Beta code per le opere in digilibLT

Se il Beta code è il passato e non ci sono al di là del TLG e del PHI altre raccolte di testi che lo adottino, sembrerebbe ovvio non preoccuparsi di marcatura secondo tale standard delle opere di digilibLT. Ma davanti a migliaia di studiosi nel mondo correntemente gestiscono e interrogano i CD-ROM del TLG e del PHI utilizzando uno o più dei programmi ad essi dedicati (LBase, TLG Workplace e PHI Workplace, SNS Greek, Musaios, Lexis, Lector, Diogenes) e poiché digilibLT tra le sua finalità si pone quella di completare l’iniziativa di digitalizzazione della letteratura latina lasciata a metà dal Packard Humanities Institute, è ovvio che il completamento della raccolta preveda un formato di marcatura identico a quello adottato per la prima parte prodotta dal PHI: è buona pratica di tutti i prodotti informatici che “continuano” qualcosa di preesistente non dimenticare la necessità della compatibilità in basso con la versione precedente. Ma c’è anche il riconoscimento del fatto che per molti studiosi che utilizzano strumenti informatici l’accesso ai dati è assolutamente preminente sull’avanzamento tecnologico dello strumento utilizzato; e se tale accesso avviene ricorrendo ad uno strumento datato (come potrebbero essere

³⁶ Un link tipizzato di tipo “conferma” che va da A verso B significa che A conferma B; mentre un link tipizzato “nega” significa che A nega B.

Musaios o Lexis, tanto per citarne due) a digilibLT interessa che lo studioso possa continuare ad accedere con tale strumento anche alle opere scaricate da digilibLT.

Pluralità di formati

Nella prospettiva di una biblioteca ad accesso aperto che contiene testi liberi da diritti, la possibilità del *download* delle opere sotto una specifica licenza d'uso (di cui si è parlato qui sopra a proposito delle licenze Creative Commons) è un servizio importante che mette il lettore e lo studioso in condizione di lavorare sul testo nel modo che preferiscono. Differenti modalità di lavoro e di studio sul testo si indirizzano verso differenti formati del file dell'opera. Partendo dal formato TEI-XML che costituisce la base della raccolta di una biblioteca digitale, altri formati si possono ricavare, e principalmente: testo puro, PDF, e-book in formato e-pub. Alcuni di essi in *digilibLT* sono già disponibili, altri sono progettati e in preparazione:



Alla base c'è l'idea che il lettore e lo studioso non possono (non devono) essere vincolati a lavorare sui testi con i programmi e gli strumenti che i curatori della biblioteca hanno predisposto per loro.

La disponibilità delle opere in formato testo puro permette di lavorare per esempio con programmi per analisi testuali, o per analisi statistiche; il formato PDF – benché proprietario – è particolarmente adatto alla stampa per tutti coloro che amano poter leggere, e annotare, su carta; il formato e-pub permette di leggere l'opera, annotarla e farvi semplici ricerche con un programma di lettura su computer o smartphone e/o con un dispositivo dedicato (e-book reader), ma soprattutto permette la creazione di una biblioteca personale che si porta sempre con sé ed è sempre consultabile, con implicazioni non ancora sufficientemente esplorate³⁷. Senza dimenticare altri formati che possono essere adatti a (o necessari per) situazioni specifiche diverse da caso a caso.

Il concetto di accesso aperto si estende quindi analogicamente a includere le modalità di studio e lettura: nessuno scienziato serio accetterebbe che chi gli fornisce i dati lo obblighi a studiarli con determinati strumenti piuttosto che altri; o meglio:

³⁷ Per ricordare due aspetti speculari nelle potenzialità e nei limiti: Petrarca portava con sé nella bisaccia nei suoi viaggi come possesso prezioso alcuni codici di piccole dimensioni; uno studioso o un lettore d'oggi possono portare con sé migliaia di libri. Però il medesimo codice che Petrarca portava con sé lo leggiamo ancora oggi; la vastissima biblioteca conservata in un dispositivo informatico tra cinquant'anni sarà certamente illeggibile.

nessuno scienziato serio accetterebbe di ricevere dati “pre-prodotti” perché il modo in cui dai fenomeni si ricavano i dati (il modo in cui i fenomeni vengono tradotti in dati) è la prima e più importante fase della ricerca. Nessuno studioso serio dovrebbe quindi accettare che un insieme di testi possa essere studiato solo con un programma realizzato dal fornitore dei testi; ma questo è proprio ciò che avviene quando si acquista l’accesso ad una raccolta di testi *on-line* o su supporto ottico. Se lo si fa deve essere chiaro che si tratta di un compromesso, non di una situazione ottimale, perché non si ha il pieno controllo sul modo in cui i testi vengono pre-lavorati³⁸.

Che la ricerca abbia necessità di basi di dati testuali aperte che permettano allo studioso di scegliere liberamente il software con cui lavorare (o perfino di crearne altro) lo scrisse in modo autorevole già alcuni fa Peter J. Heslin proprio nella recensione a una base di dati testuali aperta quale il TLG CD-ROM E:

the policy of distributing data only, and leaving users to create or purchase their own software for accessing it, is precisely the correct one for a serious scholarly project. Contrast many other databases, such as the Bibliotheca Teubneriana Latina, which can only be accessed via vendor-supplied software. This means that it cannot be peer reviewed; that you are restricted to using a certain type of computer and operating system to access the database; that you can only ask it the type of questions that have been envisioned by the designers and allowed by their software; and that you cannot easily and independently verify the correctness of the responses given by the software. In a recent BMCR review, Anne Mahoney made the point more generally: “Electronic texts are most valuable when they can be used by other scholars, not just their creators, perhaps for purposes the creators did not anticipate ...” (BMCR 01.07.10 (10 Jul 2001), Hockey, *Electronic Texts in the Humanities*, reviewed by Anne Mahoney). It is the open architecture of the TLG that makes it possible for scholars to use it creatively, and this is one of its most important features³⁹.

Benché scritte nel 2001 queste parole non hanno perso la loro rilevanza ed esprimono molto bene il punto di vista del progetto digilibLT.

Contestualizzazione dell’opera

Le opere che una biblioteca digitale distribuisce possono essere utilmente accompagnate da informazioni che aiutino il lettore a collocarle nel loro contesto, sia esso culturale, storico, espressivo ... Se pensiamo a una biblioteca digitale di testi letterari, collocare le opere nel contesto può avvenire ricorrendo a schede di presentazione e a una bibliografia di supporto.

³⁸ Con il termine pre-lavorazione si vuole indicare che alcune o tutte le occorrenze della cosiddette parole vuote possono essere eliminate; che termini fra loro sinonimi possono essere raccolti sotto un unico termine di riferimento ecc. Il fatto che la pre-lavorazione dei testi operata dalla biblioteca digitale con cui si lavora sia corretta, efficace, ottima, non elimina il problema.

³⁹ P.J. Heslin, TLG *Thesaurus Linguae Graecae, CD-ROM disk E*. Distributed by Thesaurus Linguae Graecae for the Regent of the University of California, February 2000, «Bryn Mawr Classical review (BMCR)», 23 settembre 2001, <http://ccat.sas.upenn.edu/bmcr/2001/2001-09-23.html>.

Presentazione

La pagina di presentazione dell'opera può descriverne in termini concisi il contenuto, le ragioni di interesse, la fortuna nel corso del tempo. In questo modo il lettore viene introdotto all'opera stessa; se non la conosceva e l'ha reperita esclusivamente con un criterio estrinseco quale una segnalazione o una citazione inizia a capirne qualcosa.

De hortis
Gargilius Martialis Scheda Autore

saec. III

[LEGGI IL TESTO](#)

Download possibili

[TXT](#) [TEI](#) [BETA](#) [PDF](#) [E-PUB](#) [SCHEDA CATALOGRAFICA](#)

Edizione di riferimento:

Q. Gargilii Martialis. De hortis. Introduzione, testo critico, traduzione di Innocenzo Mazzini, Bologna, Pàtron, 1988 (seconda edizione riveduta e aggiornata)

De hortis è il titolo con cui è conosciuto un breve frammento di argomento agronomico scritto da Gargilio Marziale, conservato in un palinsesto (Napoli, Bibl. Naz. IV.A.8, ff. 40r-47v), di cui costituisce la scrittura inferiore. Risale a un periodo non posteriore al V secolo. Il frammento preserva in modo incompleto e lacunoso il testo di quattro capitoli relativi ad altrettanti alberi da frutto: sulla coltivazione del cotogno, del pesco, del mandorlo e del castagno. Trattandosi di un testo acefalo, diversi sono

Operazioni preliminari: Alice Borgna
Preparazione testo: David Paniagua
Codifica TEI: David Paniagua
Edizione digitale a cura del gruppo digilibLT - Università degli Studi del Piemonte Orientale

Bibliografia

La bibliografia di supporto svolge la stessa funzione, ma in una forma differente: quando il lettore apre la pagina relativa a una determinata opera, oltre alla presentazione dell'opera gli appare una lista selezionata di testi critici che discutono dell'opera stessa. Per creare interazione fra opera antica e bibliografia critica si producono schede bibliografiche di supporto, che descrivono opere di critica riguardanti

una o più opere possedute fuori della biblioteca digitale, ma utili per lo studio di opere che si trovano al suo interno. Ogni scheda contiene uno o più campi in cui vengono indicate le opere commentate: se l'opera critica X commenta le opere D e G, quando il lettore entra nella lettura dell'opera D o G vede comparire nella bibliografia critica di supporto l'opera X. In questo modo se l'approccio del lettore è un approccio di studio la biblioteca stessa lo guida nei primi passi dello studio⁴⁰.

<p>De hortis Gargilius Martialis Scheda Autore</p> <p>saec. III</p> <p style="text-align: center;">LEGGI IL TESTO</p> <p>Download possibili</p> <p>TXT TEI BETA PDF E-PUB SCHEDA CATALOGRAFICA</p> <p>Edizione di riferimento:</p> <p>Q. Gargilii Martialis. De hortis. Introduzione, testo critico, traduzione di Innocenzo Mazzini, Bologna, Pàtron, 1988 (seconda edizione riveduta e aggiornata)</p> <p><i>De hortis</i> è il titolo con cui è conosciuto un breve frammento di argomento agronomico scritto da Gargilio Marziale, conservato in un palinsesto (Napoli, Bibl. Naz. IV.A.8, ff. 40r-47v), di cui costituisce la scrittura inferiore. Risale a un periodo non posteriore al V secolo. Il frammento preserva in modo incompleto e lacunoso il testo di quattro capitoli relativi ad altrettanti alberi da frutto: sulla coltivazione del cotogno, del pesco, del mandorlo e del castagno. Trattandosi di un testo acefalo, diversi sono</p> <p>Operazioni preliminari: Alice Borgna Preparazione testo: David Paniagua Codifica TEI: David Paniagua Edizione digitale a cura del gruppo digilibLT - Università degli Studi del Piemonte Orientale</p>	<p>Bibliografia</p> <ol style="list-style-type: none"> 1. Ferraces, A. Texto y tradición indirecta. Reflexiones sobre una edición reciente de Gargilio Marcial 2. Maire, B Gargilius Martialis, Les remèdes tirés des légumes et des fruits. Texte établi, traduit et commenté 3. Mazzini, I. Q. Gargilii Martialis De hortis
--	--

Scheda catalografica

Un terzo elemento indispensabile è la scheda catalografica che descrive l'opera. I campi che essa contiene possono essere differenti e sfaccettati a seconda dello scopo per cui la scheda è creata: uso solo interno della biblioteca digitale, oppure interazione con altre raccolte (come Internet Culturale o Europea). Al di là degli aspetti tecnici di base costituiti dall'adozione dello standard Dublin core per la scrittura della scheda, molta attenzione deve essere posta nella scelta dei campi perché essi poi permetteranno di raffinare le ricerche. Una descrizione delle opere per argomenti e generi letterari per esempio, permetterà poi di raffinare le ricerche di testo in base a questi criteri.

⁴⁰ Questo approccio, per cui la bibliografia non è generica/generale ma contestualizzata, è stato portato al progetto dai colleghi A. Balbo ed E. Malaspina che l'avevano già in precedenza adottato per il sito Web Tulliana, <http://www.tulliana.eu/home.php?LANG=I&PAG=H>.

De hortis**Gargilius Martialis**

Scheda Autore

saec. III

Scheda catalografica**ID:** DLT000233**Tipologia testuale:** Agronomia, Botanica**Citazione fonte:** Q. Gargili Martialis. De hortis. Introduzione, testo critico, traduzione di Innocenzo Mazzini, Bologna, Pàtron, 1988**Fonte BID:** CFI0142037**Edizione digitale a cura del gruppo digilibLT - Università degli Studi del Piemonte Orientale**

Presentazione dell'opera, bibliografia selezionata, scheda catalografica richiedono l'esistenza di un gruppo di studiosi, di un comitato scientifico, che provveda a un lavoro complesso di selezione e sintesi, un oscuro lavoro redazionale di reference che pone al centro dell'attenzione contemporaneamente il lettore e l'opera, l'opera che deve essere in certo modo spiegata e mostrata (anche perché il primo contatto che avviene nel mondo fisico con lo sfogliare qua e là il libro non è possibile nel mondo digitale) e il lettore che deve essere introdotto a capire se l'opera che ha individuato risponde alle sue necessità o ai suoi interessi.

Funzioni di ricerca

Le funzioni ricerca di una biblioteca digitale si collocano in due aree che in realtà poi confluiscono in un'unica interfaccia: ricerca di una specifica opera all'interno del catalogo; e ricerca di testo all'interno di una o più opere. I campi di descrizione delle opere adottati in fase di definizione della scheda bibliografica permettono di raffinare le ricerche (opere con la parola storia nel titolo quando il genere letterario è storico; oppure opere con la parola storia nel titolo quando il genere letterario è romanzo).

La ricerca di testo e l'individuazione di un'opera possono combinarsi, come accade quando si vuole individuare quali sono le opere in cui è presente un determinato argomento: in tal caso occorre individuare una o più parole che con la loro presenza segnalano con ragionevole approssimazione la presenza dell'argomento; la ricerca di tali parole sull'intera raccolta posseduta dalla biblioteca indicherà – con ragionevole approssimazione, come si è detto⁴¹ – quali sono le opere che trattano l'argomento.

⁴¹ Il primo e principale modo espressivo per cui la presenza di una parola non comporta necessariamente la presenza dell'argomento che ad essa è connesso è la negazione: chi, essendo interessato alle popolazioni che alla dissoluzione dell'impero romano entrano in Italia, cerca la parola Goti troverà anche un'ipotetica opera che contiene un'ipotetica frase «in questo libro non si

Ricerca all'interno dei testi

tutte (AND) almeno una di (OR) queste parole

Racchiudi le parole tra apici (") per cercare sequenze esatte

Parole da escludere:

Prossimità

Sezione testuale

tutto il testo titoli verso passo in greco

Criteria di selezione delle opere

Titolo

tutte almeno una di queste parole

Autore

tutte almeno una di queste parole

Tipologia testuale

Datazione

II III IV V VI VII Datazione incerta

Una comunità di studio

Oggi non è più possibile concepire un sito che costituisca un punto di riferimento per gli studiosi e non prevedere in esso lo spazio in cui gli studiosi cui ci si rivolge possano farsi sentire, interagire, scrivere. È ormai diffuso il riconoscimento del fatto che il Web permette di comunicare, di interagire, di scambiare informazioni, commenti, idee, dal basso e in orizzontale. Che i siti Web non sono necessariamente una costruzione calata dall'alto, ma che possono aprire spazi di azione e di produzione di contenuto agli utenti, e che quindi la comunità degli studiosi può essere più che una definizione ideale o astratta. Non stiamo scoprendo nulla, il mondo intorno a noi evolve e noi con lui.

parlerà dei Goti»: l'opera contiene la parola, ma non tratta l'argomento. Simmetricamente l'assenza di una parola non significa che non sia presente l'argomento: chi cercasse la parola Goti non troverebbe l'ipotetica frase «in questo libro si parlerà delle popolazioni gotiche» oppure «in questo libro si parlerà delle popolazioni chiamate Gothones dagli scrittori latini» oppure «in questo libro si parlerà delle popolazioni originarie nella regione scandinava di Götaland e poi separate nei due gruppi dei Visigoti e degli Ostrogoti». Qualche avvertenza può parzialmente ovviare al problema, che però è di portata superiore (impossibilità di prevedere tutte le frasi capaci di esprimere un metacontenuto come «questo libro parla dei Goti»).

Quindi la logica evoluzione della biblioteca digitale sarà verso la collaborazione con gli studiosi in varie forme che si possono raccogliere sotto la descrizione generale di “spazio per attività di studio e di pubblicazione”. Cioè il sito una volta avviato, a mano a mano che i singoli testi siano stati digitalizzati, corretti e codificati, offrirà agli studiosi che lo desiderino - oltre a uno strumento consueto come il forum - la possibilità di prendersi cura di uno specifico testo nel corso degli anni, di “adottarlo” per così dire, tenendo aggiornata la bibliografia che lo riguarda, descrivendo in apparato critico i punti più difficili o controversi del testo, scrivendo su di esso e pubblicando sul sito stesso (che dovrebbe dotarsi delle strutture - formali soprattutto - necessarie per gestire una rivista elettronica) o lasciando in copia sul sito un pre-print, o la versione finale, di articoli scientifici che riguardino il testo “adottato”.

Riassumendo ciò che digilibLT fa e potrà fare:

- parte dalla digitalizzazione dei testi;
- inserisce i testi in un contesto di studi critici di tipo storico, letterario ... ;
- allarga la prospettiva all’analisi testuale sui testi disponibili;
- rende possibile a chi lo desidera di scaricare i testi digitalizzati che risultino interessanti
 - in base all’analisi testuale effettuata;
 - in seguito a una scelta operata con i normali criteri di catalogazione quali nome dell’autore, titolo dell’opera, periodo, genere, forma/genere letterario, ... ;
- offre uno spazio di pubblicazione di materiali più “semplici” (come per esempio una bibliografia) o più complessi (come per esempio un articolo)
 - sia a chi abbia preso stabilmente in cura una determinata opera;
 - sia a chi voglia contribuire al di fuori di un rapporto di collaborazione stabile.

È chiaro che una prospettiva di collaborazione ampia con gli studiosi, fino a rendere loro possibile di pubblicare sul sito materiali di studio si deve confrontare con due tipi di problema:

- la validità formale di un lavoro originale pubblicato in digitale è riconosciuta, ma almeno in ambito umanistico i luoghi della pubblicazione digitale non hanno il medesimo prestigio delle riviste cartacee;
- tutti siamo studiosi, come si distingue uno studioso autorevole (“un vero studioso”) da uno studioso non autorevole?

Il primo problema ha una prospettiva lunga, digilibLT durerà nel tempo e ciò che oggi sembra remoto tra qualche anno potrà risultare affermato e diffuso: quindi si investe sul futuro. È altresì noto che i contratti di edizione con le riviste cartacee spesso vincolano gli autori, quasi sempre per pubblicare si devono cedere tutti i

diritti di riproduzione del lavoro stesso⁴², cosa che sembrerebbe escludere la possibilità di pubblicare anche su web un articolo (già) pubblicato su carta. Proprio per questo è utile valorizzare i *pre-print* e fare maggiormente leva, quando si tratta con gli editori, sulla dichiarazione di Messina sull'*open access* che molte università italiane hanno sottoscritto.

La seconda questione richiede che per pubblicare nel sito nella biblioteca (tra gli studi su un testo specifico e nella vera e propria rivista elettronica) o per le richieste di "adozione" di un testo si operi utilizzando un sistema di *referee*, inizialmente ad opera dei latinisti del progetto, e poi ad opera di un comitato scientifico di studiosi di latino tardo, così come accade di consueto per le riviste scientifiche.

La comunità degli studiosi da (ri)creare intorno ai testi

Nella storia non molto lunga dell'informatica il tema della comunità ha avuto alterne fortune. In prospettiva storica se ne può intravedere un'evoluzione che parte dal Memex inventato da Vannevar Bush negli anni Trenta (e descritto poi dieci anni dopo nel saggio *As We May Think*), concepito come una biblioteca o archivio che sono estensione della memoria individuale di uno studioso che lavora da solo:

una futura apparecchiatura per uso individuale, che sia ad un tempo una sorta di archivio e biblioteca privati. [...] Un memex è un apparecchio nel quale un individuo registra i propri libri, il proprio archivio, e le proprie comunicazioni personali [...]. È una estensione privata della sua memoria⁴³.

Proprio partendo dalla lettura del saggio di Bush, circa venticinque anni dopo Douglas Engelbart ideò con l'online System un sistema informatico per il potenziamento delle capacità intellettuali dei *knowledge workers* e per svilupparlo costituì un centro di ricerca (ARC, Augmentation Research Center) in cui egli dirigeva il gruppo di tecnici e sviluppatori con metodi e criteri aziendalistici (tanto da suscitare talora reazioni di ostilità)⁴⁴. L'utilizzatore del computer, delle risorse informatiche non è più uno studioso isolato ma un lavoratore che fa parte di un gruppo all'interno del quale avvengono scambi di informazioni.

Quando i tempi portano allo sviluppo del personal computer si è di nuovo di fronte a un dispositivo posto sotto il segno dell'individuale, come il nome stesso indica. Ma già nel 1968 Robert Taylor in visita all'ARC, il centro di ricerca di Engelbart, aveva visto i prototipi di ciò che sarebbe poi diventato al PARC della Xerox il personal computer e aveva raccontato la visita e le riflessioni che essa gli aveva susci-

⁴² Non mancano situazioni meno vincolanti di quella qui delineata, ma sono casi che non riescono ancora a definire una chiara tendenza o un definito orientamento in atto.

⁴³ Vannevar Bush, *As We May Think*, «The Atlantic Monthly», 176 (1945) n.1, p.101-108.

⁴⁴ Thierry Bardini, *Bootstrapping: Douglas Engelbart, coevolution, and the origins of personal computing*, Stanford: Stanford University Press, 2001. Il tema sottende l'intero cap. 6.

tato, in un articolo intitolato *The computer as communication device*⁴⁵. In altre parole nello strumento informatico personale, individuale, erano presenti caratteristiche che agli occhi di Taylor erano di gran lunga più importanti, quelle che lo rendevano adatto per comunicare: cioè per scambiare informazioni, per stabilire relazioni. Oggi per chi non è un professionista dell'informatica è consueto utilizzare il personal computer come dispositivo di comunicazione, e non solo di elaborazione dati in senso riduttivo⁴⁶, verso persone o mondi con i quali si hanno relazioni, si condividono dati e informazioni, si scambiano opinioni, si formulano nuove idee. Questi cenni, incompleti ma significativi, indicano un movimento della finalizzazione degli strumenti digitali che va dall'individuale al comunitario (senza idolatrare ciò che si può collocare nella categoria di comunitario né svilire ciò che si può mettere nella categoria di individuale). Il movimento complesso, alterno, ma indirizzato, verso la collaborazione tra studiosi può costituire il catalizzatore capace di far scaturire dalle biblioteche digitali approcci e interpretazioni innovative basate su una visione completa dei testi di una lingua e di una civiltà, e non solo su una visione settoriale spesso centrata su un singolo testo. L'indagine centrata su un singolo testo, per quanto valida, rimane legata a quella fase degli studi in cui lavorando su carta ciò era già molto impegnativo. Oggi si può fare altro, occorre pensare come farlo, e poi farlo.

Today a digital library is no more by itself a groundbreaking product, rather it is a daily tool for the work of scholars, students, librarians so it must as much as possible answer in the best way to a lot of needs and requirements.

digilibLT (digital library of late-antique latin texts; <http://digiliblt.unipmn.it>) is a newly born digital library which collects all the late Latin pagan texts, as a complement to the collection of classical Latin texts known as PHI cdrom #5.3. The library has been designed following the announcement of a competitive examination of research proposals made by Regione Piemonte. The aim is to answer the main needs of scholars and librarians. Ancient Latin texts are - or will be soon - available in various formats from XML-TEI to texts, PDF and epub (as well as the old Beta coding used by PHI for the texts of its 5.3 cdrom); they all come from authoritative recent critical editions which are selected by the scientific committee of the library. Introductions, comments, apparatuses contained in those same editions are completely absent as requested by the law about intellectual property.

The whole system is built with free software and adopts the principles of Open Access and Creative Commons licenses for all of its contents (web pages and texts).

⁴⁵ Robert Taylor, *The computer as a communication device*, «Science and Technology», 1968, n.76, <http://www.cs.utexas.edu/users/chris/sigcomm/t1/liclider.pdf>. Se è oggi ovvio pensare il computer come un dispositivo di comunicazione, può aiutare a cogliere il paradosso del suo titolo tradurlo con qualcosa come «la calcolatrice come dispositivo di comunicazione».

⁴⁶ È ovvio che anche inviare un'e-mail implica elaborazione di dati da parte del processore.

* Per tutti i siti web l'ultima consultazione è avvenuta il 30 giugno 2012.