

Le sfide di una biblioteca digitale del latino tardoantico

Struttura, canone e questioni aperte di codifica e visualizzazione

Alice Borgna

(Università del Piemonte Orientale, Italia)

Simona Musso

(Università del Piemonte Orientale, Italia)

Abstract This paper focuses on different aspects of the *DigilibLT* project (Digital Library of Late Antique Latin Text), illustrating not only its functionality, but also problems and difficulties in terms of project building and management. To create a digital library, the indisputable starting point is the definition of the canon of the texts to be included; in this sense, our selection, based on late antique Latin texts, pagan and in prose, revealed some critical point. First of all, since most of these works have technical-scientific or educational purposes, they are in many cases accompanied by diagrams, images and symbols, a graphic apparatus that often can be made digital only with some difficulties. Besides, on a more general level, we will also explain how the strong participation of young people in the team has played a crucial role in the development of the project; it has also contributed to enrich the CV of plenty humanist graduates with significant competencies in the field of the Digital Humanities.

Sommario 1 Costruire una biblioteca digitale. – 1.1 Consultazione e *download*. – 1.2 La ricerca. – 1.3 Un ambiente di lavoro: schede e bibliografia. – 2 Canone. – 3 Formati e visualizzazione. – 4 Glifi. – 5 Natura collaborativa e qualità: un rapporto problematico. – 6 Digitale, perché? – 6.1 Un punto di partenza.

Keywords Digital Libraries. Project building. XML-TEI encoding. Glyphs. Late antiquity.

La rivoluzione informatica che ha investito tutti i campi di studio, anche quelli che potevano apparire maggiormente legati a strumenti tradizionali, come la filologia classica, negli ultimi decenni sta conoscendo un'evoluzione altrettanto importante. Molte delle banche dati testuali nate su CD-ROM, infatti, oggi più o meno ovunque vengono riversate sul web, spesso

con qualche difficoltà,¹ a dimostrazione di un problema, quello della rapida obsolescenza del supporto tecnologico, che rappresenta una delle grandi sfide della contemporaneità. Non solo: la dimensione online di database e biblioteche digitali suscita anche questioni sostanziali, legate all'accesso libero e pone in primo piano la necessità che risorse affini trovino il modo di coordinarsi e collaborare.

Questi rapidi sviluppi hanno anche reso urgente la sistemazione digitale di quel materiale che ne è ancora privo, una lacuna che può condizionare fortemente anche lo stesso progredire della ricerca scientifica. Era questo il caso della letteratura latina tardoantica che, fino a qualche anno fa, la rete offriva in forme sporadiche e non organizzate, ovvio riflesso di una minor fortuna, anche a livello cartaceo, dell'approfondimento critico in merito alle opere dei secoli II-VI d.C.² Con l'intento di ovviare a questa assenza, nel 2010 è nato il progetto *Digital Library of Late Antique Latin Texts* (digilibLT), che ha costruito e via via implementato una biblioteca digitale dedicata in modo particolare alla prosa pagana.³

1 Costruire una biblioteca digitale

1.1 Consultazione e download

Premessa necessaria alla costruzione di una risorsa digitale è stata, naturalmente, una prima stesura di un canone di riferimento che permettesse l'individuazione di tutte le opere e di tutti gli autori da comprendere.⁴ Questo canone, di circa quattrocento voci, non è solo scaricabile come

1 È questo ad esempio il caso, su cui vi sarà modo di tornare, del prototipo di CD-ROM (*GL-CD*) contenente il corpus delle opere grammaticali. Cf. *infra* § 3.

2 Lo stesso *Thesaurus linguae Latinae*, con il quale *DigilibLT* ha avuto modo di comunicare (cf. *infra* § 6), non offre una schedatura completa delle opere dalla metà del II secolo in poi, soprattutto quelle di contenuto profano. Per lo stato frammentario e disomogeneo in cui la rete offriva i testi della letteratura latina tardoantica prima di *DigilibLT*, cf. Tabacco 2014, 3-16, Borgna 2017, 4-5.

3 URL www.digiliblt.uniupo.it. Il progetto è stato finanziato dalla Regione Piemonte in seguito a un bando competitivo finalizzato a ricerche nell'ambito delle Scienze umane e sociali. Il gruppo di ricerca, che fa capo all'Università del Piemonte Orientale, è diretto da Raffaella Tabacco, responsabile della ricerca e condirettore per la parte relativa alle competenze disciplinari di Letteratura latina, Maurizio Lana, a capo delle competenze di *Digital Humanities*. L'Università di Torino è stata proponente aggiuntivo e ha visto l'installazione di una seconda unità di ricerca presso il Dipartimento di Studi Umanistici con Ermanno Malaspina e Andrea Balbo. Per i molteplici aspetti del progetto, anche informatici, si vedano Lana 2011, 2012b, 2012c; Tabacco 2014, Mosca-Musso 2016; Borgna 2017.

4 Cf. *infra* § 3.

file autonomo, ma è anche attivamente ricercabile nel sito per autore, per opera, per ordine alfabetico o per secolo. Si riporta, come esempio, la schermata che si ottiene ricercando gli autori che iniziano con la lettera P: nell'elenco compaiono, insieme con gli autori, anche le opere anonime (in corsivo).

The screenshot shows the digilibLT website interface. At the top, there is a search bar and navigation links. The main content area is titled "Opere che iniziano con la lettera P". It lists 20 items, each with a number, a set of icons (representing text availability and download options), and the title of the work. The items are:

1. *Panegyrici Latini* - saec. III-IV
2. *Panegyricus dictus Augustodunus a. 297-8* - oratio pro restaurandis scholis - saec. III
3. *Panegyricus dictus Constantino filio Constanti* (Treueris a. 313 p. Chr.) - saec. IV
4. *Panegyricus dictus Constantino imperatori* (Ramae a. 321 p. Chr.) - saec. IV
5. *Panegyricus dictus Constantino imperatori* (Treueris a. 310 p. Chr.) - saec. IV
6. *Panegyricus dictus Constantino imperatori* (Treueris a. 312 p. Chr.) - *Gratianus actio* - saec. IV
7. *Panegyricus dictus Constantio Caesari* (Treueris a. 297-8) - saec. III
8. *Panegyricus dictus Iuliano imperatori* (Constantinopolis a. 362 p. Chr.) - *Gratianus actio de consulatu suo* - IV saec.
9. *Panegyricus dictus Maximiano et Constantino* (Treueris 7 a. 307 p. Chr.) - saec. IV
10. *Panegyricus dictus Maximiano et Diocletiano* (Treueris a. 285 p. Chr.) - saec. III
11. *Panegyricus dictus Theodosio imperatori* (Roma a. 389 p. Chr.) - saec. IV
12. *Panegyricus dictus Treueris 7 a. 291 p. Chr.* - *Genethlacus Maximian Augusti* - saec. III
13. *Partitiones* - saec. V-VI
14. *Passio Sanctarum Perpetuae et Felicitatis* - 203 ca.
15. *Pauli Diaconi Epitome* - saec. II (VII)
16. *Peripigratio Egeriae* - 384 ca.
17. *Peri hermeneias liber Aristotelis latine versus* - saec. VI
18. *Periochae Lulli* - saec. III-IV
19. *Philosophiae consolatatio* - 523-524
20. *Physiologus latinus rec. II* - saec. IV-V

Figura 1. Autori e opere anonimi che iniziano con la lettera P

Ancora più ricca di informazioni è la ricerca per opera (fig. 2), che segna con apposite icone la presenza del testo all'interno della banca dati, precisando in quale formato essa sia disponibile. I testi, infatti, oltre ad consultabili online, sono anche scaricabili gratuitamente (previa una registrazione a soli fini statistici) e per intero in vari formati (.txt, TEI, PDF,

ePub), in modo che ogni studioso possa utilizzarli nel modo più adeguato ai suoi interessi di ricerca o di lettura.⁵ Tutte le opere sono poi distribuite su licenza Creative Commons (cf. Lana 2012a, 48-64), nel rispetto della filosofia *open source* e in modo da rendere il lavoro già fatto la base di partenza per altra ricerca (cf. *infra* § 7.2).



Figura 2. Situazione al 2015-06-25

Un elenco preciso di tutte le opere disponibili per la ricerca e il download (divise per secoli) al momento in cui si consulta la biblioteca digitale è fornito dal primo pulsante sulla colonna di sinistra. Ogni nuova accessione è poi segnalata in home page da un'apposita notizia. La figura 3 riporta, come esempio, l'elenco delle opere del V secolo a disposizione dell'utente.

5 Sulle ragioni per cui si è deciso di abbandonare il formato Beta Code, ormai destinato all'obsolescenza, cf. Tabacco 2014, 10 n. 21.

The screenshot shows the digilibLT website interface. At the top, the logo 'digilibLT' is displayed alongside the text 'Biblioteca digitale di testi latini tardoantichi'. To the right are logos for 'UPO' and 'REGIONE PIEMONTE'. Below the header, there is a navigation bar with links like 'Home', 'Il progetto', 'Notizie', 'Tardoantico nel web', 'Aiuto', 'Contatti e feedback', and 'Accedi'. The main content area is divided into a left sidebar and a main list.

Ricerca

Opere disponibili per la ricerca
 Ricerca avanzata
 Ricerca bibliografica

Letture e download di testi

Opere
 Autori
 Bibliografia

Per ordine cronologico
 e il IV V VI VII VIII
 Datazione incerta
 Bianco opere

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

Ricerca

Opere sul tardoantico
 Canone autori tardoantichi
 Download di fonti e software
 Scarica tutte le opere

Propaganda aggiuntiva

Scegliere nella parte di sinistra l'ordinamento:
 Cronologico (per secolo, datazione incerta o elenco completo)
 Alfabetico (per lettera iniziale o elenco completo)

Opere V secolo

1. Adnotationes super Lucanum
2. Adnotationes super Lucanum: supplementum
3. Albercatio Hadriani Augusti et Epictati philosophi
4. Anecdota medicum
5. Ars (sive de nomine et serbo)
6. Ars Bobemisa (= Excerpta Bobemisa GL 1 533-565)
7. Ars Breuiata
8. Ars de orthographia
9. Ars grammatica (Asper minor)
10. Ars grammatica (Cledonius)
11. Ars grammatica (Pl. Asper)
12. Brevis expositio Vergili georgicorum
13. Celeres passionis
14. Collatio Alexandri et Diodini
15. Commenta Bernensia in Lucanum
16. Commentaria in metra Terentiana
17. Commentarii in dictiones in Quintum Caecilium
18. Commentarii in Verrinas
19. Commentarium de oratione et de octo partibus orationis artis secundae Donati
20. Commentarius in Vergili Aeneidos libros

1 2 3 4 5 6 »

Figura 3. Elenco delle opere del V secolo a disposizione dell'utente

Oltre a questi testi è poi possibile scaricare una serie di opere di critica sul tardoantico di cui si è prodotta un'edizione anastatica digitale in formato PDF e ricercabile per parola: si tratta di saggi liberi dal vincolo di copyright, oppure per cui si è ottenuto il consenso alla distribuzione da parte dei titolari.⁶

⁶ Punto di partenza sono state le opere di Italo Lana, tra i primi latinisti italiani a porre speciale attenzione al tardo antico e nel cui magistero si sono formati molti dei membri del progetto: un buon numero di questi studi, nati come dispense universitarie, non solo oggi possono essere reperiti con difficoltà, ma la loro stessa sopravvivenza è messa in pericolo dal rapido processo di disgregazione della carta acida su cui sono stati stampati.

1.2 La ricerca

All'utente della biblioteca si aprono due porte di accesso alla ricerca: una semplice e una avanzata. Nel primo caso è possibile interrogare il corpus mediante il semplice inserimento di una parola nella home page; le forme flesse si ottengono inserendo la radice della parola seguita da *. La marcatura a cui sono stati sottoposti i testi⁷ consente però anche interrogazioni più raffinate, possibili grazie alla maschera di ricerca avanzata: sono qui possibili ricerche non solo sull'universo dei testi, ma anche su sottosezioni stabilite dall'utente (opere singole, tipologie testuali specifiche, per secolo, ecc.). Asterischi e operatori booleani permettono di combinare le ricerche di più parole con desinenze variabili; determinabile è anche la prossimità delle parole oggetto di ricerca e la sezione testuale specifica in cui sono contenute (nell'intero testo, nei titoli, nei versi o nei passi in lingua greca).



Figura 4. Ricerca all'interno dei testi

7 Per una descrizione dettagliata delle scelte di marcatura del progetto cf. *infra* §4, con ulteriore bibliografia.

I risultati della ricerca danno accesso alle occorrenze dei termini cercati nelle singole opere, ciascuna voce permette poi l'accesso al testo completo dell'opera e a una concordanza (KWIC, Key Word In Context), visualizzabile secondo l'ordine del testo, per contesto sinistro o per contesto destro. L'immagine che segue (fig. 5) mostra la concordanza, in ordine del testo, delle ricorrenze di *pax* nell'*Epitoma historiarum Philippicarum Pompei Trogi* di Giustino.

The screenshot shows the digilibLT website interface. At the top, there is a search bar with the text 'Cerca' and a 'GO' button. Below the search bar, there are several navigation and utility links: 'Home', 'Il progetto', 'Notizie', 'Fornimento del web', 'Aiuto', 'Contatti e feedback', 'Accessi'. On the right side of the header, there are logos for 'digilibLT', 'Biblioteca digitale di testi latini tardoantichi', 'UPO', and 'REGIONE PIEMONTE'. The main content area displays the search results for the word 'pax' in the text 'Iustinus, Epitoma historiarum Philippicarum Pompei Trogi'. The search results are organized into a table with columns for the search term, the context (left and right), and the page number. The search term is 'pax' and the context is 'KWIC'. The search results are as follows:

Ordinamento	Contesto sinistro	Contesto destro
x. 5.8	...paxem suscepit ubi fuerunt. Post haec	...ad dicitur fuit angue tempor...
v. 10.4	...ad inferos subit. Quibus nihil curi	...statu ubi, mirabile dicitur re...
vi. 4.8	...ad tantum regni fuit, quod re...	...ista Graecia tanta est, qui man...
vii. 4.2	...mibi ubi ex comitiis utrumque	...tanta. Ex inferis quique Graecia...
ix. 2.7	...identem recuperant. Igitur non	...tantum Agrippae concipitur, ubi...
xviii. 2.16	...antem si qui Graecum potest fieri. Sed	...ad hunc illa memoratur a Roma...
xxv. 5.16	...ad multas curas rebusque de...	...sunt et Rhodensium non convenit...
xxxiii. 4.1	...suntque. Deinde et cum Romano ex	...indicia sunt ubi et resistunt...

On the left side of the page, there are several navigation and utility links: 'Opere disponibili per la ricerca', 'Ricerca avanzata', 'Ricerca bibliografica', 'Lettura e download di testi', 'Opere', 'Autori', 'Bibliografia', 'Per ordine cronologico', 'Per ordine alfabetico', 'Ricerca', 'Opere sul tardoantico', 'Database autori tardoantichi', 'Download di font e software', 'Download testi', and 'Proponente aggiornato'.

Figura 5. KWIC *pax* in Iustinus, *Epitoma historiarum Philippicarum Pompei Trogi*

1.3 Il contesto di supporto

La definizione di 'biblioteca' (*Library*) di DigilibLT sta a indicare una natura che non vuole limitarsi al database, ma rappresentare un ambiente di lavoro dedicato al tardo antico, all'interno del quale anche il semplice interessato possa trovare un orientamento entro un panorama letterario relativamente poco noto. Per questo motivo i testi si presentano accompagnati da schede che ne illustrano i contenuti e i principali problemi critici (scheda opera) così come i dati noti sull'autore (scheda autore), un apparato spesso a firma dei maggiori esperti dei testi in questione.



Figura 6. Scheda di Giovanni Polara del *De comoedia vel de fabula* di Evanzio

La biblioteca comprende anche una ricca bibliografia, collocata sia a lato delle pagine che contengono le schede, sia accessibile dalla home page, dove un pulsante conduce a una maschera di ricerca avanzata.⁸ La figura 7 mostra il risultato di una ricerca su *Eumenius* in ordine alfabetico per autore moderno.

8 L'interrogazione è possibile attraverso diverse chiavi: autore moderno, autore antico, titolo, opera antica, tipologia editoriale (edizione, lessico, miscellanea, monografia), titolo di rivista, parole chiave descrittive del contenuto, anno di edizione, cronologia dell'autore e dell'opera.

The screenshot shows the digilibLT website interface. At the top, the logo 'digilibLT' is displayed next to the text 'Biblioteca digitale di testi latini tardoantichi'. To the right, the logos for 'UPO' and 'REGIONE PIEMONTE' are visible. Below the header, there is a navigation bar with links like 'Home', 'Il progetto', 'Notizie', 'Tardoantico nel web', 'Aiuto', 'Contatti e feedback', and 'Accessi'. The main content area is titled 'Ricerca bibliografica' and shows search results for 'Eumenius'. The results are ordered alphabetically by modern author, as indicated by the 'Ordina per' dropdown menu. The first four results are listed:

1. Carrozza F., Letteratura Latina Tardoantica. *Bibliografia riferita all'autore antico*:
1. Eumenius saec. III
2. Cavallo G., Fedeli P., Giardina A., Lo Spazio Letterario di Roma Antica. Volume II: la circolazione del testo. *Bibliografia riferita all'autore antico*:
1. Eumenius saec. III
3. Giardina A., - Silvestri M., Il principe e il testo. *Bibliografia riferita all'autore antico*:
1. Eumenius saec. III
4. Lana I., e Maltese E. V., Storia della civiltà letteraria greca e latina. Volume terzo: dall'età degli Antonini alla fine del mondo antico. *Bibliografia riferita all'autore antico*:
1. Eumenius saec. III

The interface also includes a search bar, a 'CERCA' button, and various filters and sorting options.

Figura 7. Risultato di una ricerca su *Eumenius* in ordine alfabetico per autore moderno

2 Canone

Scelta preliminare per la creazione di una biblioteca digitale è la definizione del canone delle opere in essa contenute, con la relativa scelta dell'edizione di riferimento.

Sono necessarie alcune precisazioni. La biblioteca *digilibLT* comprende i testi in prosa dal II al VI secolo d.C.; è utile ribadire che siamo partiti da zero, senza un corpus digitalizzato pregresso (con parziale eccezione per quanto riguarda i grammatici su cui torneremo). Ciò ha implicato, in primo luogo, una selezione delle tipologie di testi da includere, in modo che il progetto fosse sostenibile per le forze a disposizione. Si è quindi rinviato l'inserimento di testi giuridici e cristiani, tenendo conto che per queste tipologie testuali esistono già risorse digitali (cf. *supra* nota 2).⁹

⁹ Come si è illustrato nel § 1; cf. anche Tabacco 2014, 3-6.

Inoltre non era disponibile un canone di riferimento esauriente né la sua composizione si è rivelata un'operazione semplice. Non vi sono comprese, infatti, solo opere di alto valore letterario, si pensi alla grande storiografia di Ammiano Marcellino, ma anche breviaristica, storia romanizzata, testi tecnici, manuali scolastici, *scholia*, commenti, testi geografici, antiquari, ecc., un vasto insieme di opere eterogenee e non facilmente classificabili. Solo alcune di esse erano già disponibili online, ma non inserite in una biblioteca esauriente e corredata di strumenti che permettessero una ricerca avanzata. Spesso, poi, i testi reperibili non garantivano correttezza e scientificità o non dichiaravano neppure l'edizione critica di riferimento. E infine, ed è un punto di non secondaria importanza, i siti esistenti, pur consentendo la libera consultazione unita a funzionalità di ricerca più o meno avanzate, generalmente non prevedono di offrire il download dell'opera completa.

Possiamo quindi affermare che la biblioteca *digilibLT* è stata creata ex novo per quanto riguarda sia la scelta del materiale da inserirvi sia la preparazione di testi digitalizzati.

Per quanto concettualmente la definizione del canone preceda la digitalizzazione delle opere, in realtà si è proceduto di pari passo, scansionando, correggendo e marcando i testi man mano che venivano messe a punto le varie voci del canone. Né sono mancate ridefinizioni delle voci del canone e sostituzioni delle edizioni di riferimento per adeguarsi, ad esempio, alle pubblicazioni scientifiche avvenute dal 2010, anno di inizio del progetto, a oggi. Ma procediamo con ordine.

La biblioteca *digilibLT* presenta in realtà un doppio canone. Online è disponibile l'elenco, consultabile per ordine cronologico o alfabetico, degli autori e delle opere. Tale elenco è stato preparato da Andrea Balbo e si è visto (§ 2.1) come le singole voci consentano di accedere alla lettura dei testi e delle relative schede opera e autore. Un consistente nucleo di tale elenco era già disponibile alla messa online della biblioteca, ma nel corso degli anni è stato continuamente aggiornato e ampliato, in base al progredire degli studi e parallelamente alla preparazione del canone vero e proprio, liberamente scaricabile in formato PDF, nel quale per ogni opera è indicata l'edizione di riferimento.¹⁰

Punto di partenza per l'individuazione sia delle opere sia dell'edizione di riferimento, è stato l'*Index* del *Thesaurus linguae Latinae*,¹¹ opera fondamentale per chiunque si accinga allo studio di testi dell'antichità latina. E tuttavia il Thesaurus fornisce un quadro di tutte le opere latine, rendendo

¹⁰ Il canone è stato preparato da Simona Musso, con aggiornamenti successivi a cura del gruppo di ricerca.

¹¹ La parte del volume *Praemonenda de rationibus et usu operis* (pubblicato nel 1990) dedicata all'indice è dovuta a D. Krömer.

complessa la consultazione a chi sia interessato in maniera specifica alle opere profane tarde, né è frequentemente aggiornato.¹² Per queste ragioni il canone realizzato nel progetto, oltre a svolgere la funzione strumentale di indicare le edizioni adottate, si presenta come utile strumento di lavoro per studiosi di tarda antichità. Rispetto all'*Index* del *Thesaurus* non si è operato solo un criterio selettivo, ma sono anche stati apportati aggiornamenti e modifiche, per i quali sono state determinanti, oltre agli studi condotti, le segnalazioni di studiosi di ogni parte del mondo che hanno permesso di rimediare a inevitabili imprecisioni e omissioni. In alcuni casi, ad esempio, sono state indicate opere che secondo la datazione canonica non rientrerebbero nel periodo preso in considerazione dal progetto, ma che sono state recentemente retrodatate¹³ oppure opere la cui redazione a noi giunta è di età medievale, ma che contengono nuclei di probabile datazione tardoantica.¹⁴ Altre volte, invece, avevamo ommesso opere rimaste inedite fino ad anni molto recenti o per le quali esistono solo edizioni risalenti al XVI secolo.¹⁵

La scelta dell'edizione di riferimento si basa sul criterio scientifico dell'individuazione della migliore esistente. Il panorama molto variegato delle tipologie testuali incluse a canone produce situazioni molto diverse rispetto alla disponibilità di edizioni. Per molte opere esiste una sola edizione, altre sono state edite solo in epoca prescientifica, mentre le opere più studiate sono leggibili in molte edizioni, alcune anche molto recenti, e non sono rari i casi in cui si possa usufruire di più edizioni di alto valore. Si è quindi operata una scelta che non intende, tuttavia, essere esclusiva. In una prospettiva futura, sarà possibile aggiungere l'apparato critico ai testi digitalizzati oppure affiancare più edizioni. Nel frattempo si è tenuto conto delle varie edizioni esistenti procedendo a un confronto costante, anzitutto per correggere eventuali errori di stampa: per citare un caso, si veda il *Breviarium* di Festo per il quale l'edizione di riferimento è Eadie (1967), ma si è tenuto conto delle correzioni tipografiche apportate nell'edizione Fele (2009).

Discorso a parte meritano i grammatici, per i quali il progetto si avvale dei testi che vennero digitalizzati all'Università di Torino sotto la direzione scientifica di Nino Marinone. Il prototipo di CD-Rom, che rappresentò l'esito del progetto (GL-CD), tuttavia, non venne mai commercializzato, e dopo poco più di due decenni mostra già la sua obsolescenza rendendo

12 Risalgono al 2010 gli *Addenda* all'*Index* dalla Bayerische Akademie der Wissenschaften e sono disponibili online (<http://www.thesaurus.badw.de/addenda.htm>).

13 Il progresso degli studi può aver modificato non solo la datazione, ma anche l'attribuzione di un'opera a un determinato autore.

14 Un caso esemplificativo è rappresentato dagli scolii.

15 Si pensi, ad esempio, alla versione latina dell'opera medica di Alexander Trallianus.

assai complessa l'estrazione dei dati in esso contenuti e la loro riconversione per uniformarli agli standard oggi in uso e garantirne l'accessibilità. Il *GL-CD* non utilizzava, infatti, la codifica Unicode UTF-8 per i caratteri e aveva adottato un font specifico per il greco (Greco), creato ad hoc (cf. Marinone 1979; Lomanto 1980). Ai fini dell'inserimento dei testi di *GL-CD* all'interno di *DigilibLT*, a essi è poi stata applicata una marcatura *TEI/XML*, coerente con le scelte della biblioteca, in modo da consentire una ricerca avanzata sul corpus.¹⁶

Non è ora necessario entrare nello specifico della codifica dei grammatici, ma il caso in questione ci offre il destro per sottolineare l'importanza della scelta del tipo di codifica dei testi al fine di garantirne la sopravvivenza. Il ritmo vorticoso di evoluzione degli strumenti digitali rende obsolete in pochissimi anni tecnologie che oggi ci paiono all'avanguardia. L'ambizione di progetti di digitalizzazione di testi è invece quella di durare nel tempo, ma ciò è possibile solo se si svincola il documento da hardware e software specifici o da rigide visualizzazioni predefinite, si adottano standard condivisi e si distribuiscono i testi sotto licenza *Creative Commons* e non protetti e limitati da copyright.¹⁷

Limitandoci alla trattazione dei grammatici relativamente al canone, bisogna spiegare che, partendo da un corpus pregresso, si è scelto inizialmente di adottare le stesse edizioni utilizzate nel *GL-CD*, procedendo poi, laddove esistano edizioni più recenti, ad aggiornare il corpus con l'edizione di riferimento attuale.¹⁸ La suddivisione delle opere e i relativi titoli adottati nel *GL-CD*, tuttavia, spesso non coincidono con quelli adottati dall'*Index* del *Thesaurus*, soprattutto per quanto riguarda i trattatelli anonimi e i frammenti. Problema accentuato, inoltre, dall'ambiguità dei titoli generici delle opere grammaticali (*Ars*, *De metris*, *De uerbo*, ecc., per fare solo alcuni esempi).

Per rendere univoca l'identificazione dei testi, è stato quindi necessario indicare a canone non solo l'edizione di riferimento (posta in grassetto), ma anche il volume e le pagine dell'edizione Keil, aggiungendo inoltre segnalazione di eventuali edizioni più recenti rispetto a quella adottata. Portiamo ad esempio il caso degli *Exempla Elocutionum* di Arusiano Messio.

16 La conversione in Unicode e la marcatura XML sono state operate da Piero Colombo-Step Srl. Data la frequente presenza di greco e di simboli tecnici (si pensi ai simboli metrici), al fine di garantire la massima correttezza, Ermanno Malaspina sta provvedendo a un controllo linguistico, correggendo manualmente eventuali inesattezze di transcodifica. Simona Musso, invece, si sta occupando della verifica della marcatura *TEI/XML*.

17 Per le licenze *Creative Commons* si veda il sito <http://creativecommons.org/>.

18 L'elenco delle opere del *GL-CD* e delle relative edizioni è contenuto in Lomanto 1980. Per maggiori dettagli sugli studi pubblicati a partire dal 1979 sul progetto di digitalizzazione dei grammatici cf. Tabacco 2014, 5 n. 8 e Borgna 2017, 4 n. 8.

Autore	Opera	Edizione di riferimento
Arusianus Messius saec. IV ex.	<i>Exempla elocutionum</i>	<p>Arusiani Messii <i>Exempla elocutionum</i>, Introduzione, testo critico e note a cura di A. Di Stefano, Hildesheim 2011 (Collectanea Grammatica Latina 6). s. Arusianus Messius, <i>Exempla elocutionum</i>, a cura di A. Della Casa, Milano 1977</p> <p>Cfr GL VII 449-514 <i>Grammatici Latini, VII. Scriptores de orthographia, Terentius Scaurus, Veilius Longus etc.</i>, ex recensione H. Keilii, Hildesheim 1961,449-514 (reprografischer Nachdruck der Ausgabe Leipzig 1880).</p>

Figura 8. Voce del canone

Nella colonna di destra si vede in grassetto l'edizione di riferimento (Di Stefano 2011). Seguono l'indicazione dell'edizione digitalizzata nel GL-CD (Della Casa 1977) e il riferimento al volume e alle pagine dell'edizione Keil.

3 Formati e visualizzazione

L'indicazione dell'edizione di riferimento adottata non si trova solo nel canone, ma è riportata sia nella scheda opera online sia nel TEI Header dei testi digitalizzati, nel marcatore <sourceDesc> contenuto in <fileDesc>, come mostra l'esempio sotto riportato.¹⁹

19 I file TEI/XML sono suddivisi in due sezioni: il TEI Header - la sezione introduttiva in cui sono contenute informazioni relative all'opera, all'autore, alla fonte, alle modifiche apportate al testo - e il testo vero e proprio. I file del progetto *digilibLT* hanno il TEI Header suddiviso in quattro sezioni di cui l'ultima è opzionale. Nella prima sezione (<fileDesc>) sono indicati il titolo e l'autore dell'opera antica; il nome del responsabile della correzione linguistica e della marcatura (spesso non coincidenti); editore, luogo e data di pubblicazione; licenza *Creative Commons* sotto cui sono distribuiti i testi; il numero identificativo dell'opera e la citazione bibliografica dell'edizione di riferimento. La seconda sezione, <encodingDesc>, è a sua volta suddivisa in <projectDesc>, <editorialDecl>, <charDecl>. La parte più corposa è l'<editorialDecl> in cui si specificano sia le convenzioni seguite nella trascrizione e nella marcatura dei testi - distinguendo quelle generali da quelle specifiche di una singola opera - sia eventuali modifiche apportate al testo di riferimento. Il <projectDesc> riporta il nome esteso del progetto e dei suoi responsabili. Parleremo del <charDecl> affrontando la questione dei glifi. La terza sezione <profileDesc> contiene il <langUsage> in cui è specificata la lingua in cui sono scritti i testi del progetto *digilibLT*, cioè il latino. Il greco non è presente in tutti i testi e, laddove compare, costituisce generalmente una porzione nettamente minoritaria. È infine prevista la sezione <revisionDesc> in cui si può tener memoria di eventuali interventi sul file in tempi successivi alla prima pubblicazione. Per una descrizione più approfondita, cf. Musso 2014, 171-5.

```
<sourceDesc>
<p>
<hi rend="italic">Le premier mythographe du Vatican</hi>, texte éta-
bli par Nevio Zorzetti et traduit par Jacques Berlioz, Paris 1995 (Collec-
tion des Universités de France)
</p>
</sourceDesc>
```

Attraverso fogli di stile XSLT, dall'XML si ricavano i vari formati (HTML, .txt, ePub, PDF)²⁰ in ognuno dei quali compare la dichiarazione dell'edizione di riferimento. La scelta di *digilibLT* di rendere libero e gratuito il download dei testi, per intero²¹ e con la possibilità di scelta del formato più adatto alle proprie esigenze, rende la biblioteca unica nel suo genere.

Bisogna precisare che la scelta di ricavare formati con caratteristiche molto diverse fra loro da un unico file sorgente TEI/XML ha richiesto di riflettere con attenzione anche sulle possibilità di visualizzazioni. È ben noto che la marcatura TEI/XML ha un suo punto di forza proprio nel non essere vincolata a una specifica visualizzazione. La TEI, per quanto preveda anche marcatori indicanti semplicemente una formattazione,²² si caratterizza per marcatori descrittivi che determinano, ad esempio, il fatto che una porzione testuale sia un titolo, un'integrazione o una citazione e non come essa vada visualizzata. Si è già accennato che a partire dalla marcatura TEI si possono poi ottenere le più svariate visualizzazioni attraverso i fogli di stile, tuttavia ci sono alcuni casi specifici in cui determinate scelte di marcatura possono porre alcune limitazione nella visualizzazione.

Se ne può individuare un esempio nel caso in cui con il testo digitale si voglia riprodurre sia la scansione in libri, capitoli e paragrafi sia la suddivisione logica in capoversi dell'edizione di riferimento. I problemi nascono, infatti, non solo quando capitoli e paragrafi sono numerazioni indipendenti tra loro, ma anche quando uno dei due livelli strutturali (o a volte entrambi) non coincide con i capoversi, eventualità per nulla rare nei testi latini.

20 La natura dichiarativa e descrittiva di TEI/XML predica una caratteristica per una determinata porzione di testo (specificando, ad esempio, che si tratta di un titolo di paragrafo), ma non determina un output specifico per tale caratteristica (ad esempio, la sua visualizzazione in grassetto). Le opzioni di visualizzazione dei fenomeni marcati e i formati di output si ottengono attraverso fogli di stile XSLT (*Extensible Stylesheet Language-Transformation*); per la biblioteca *digilibLT* i fogli di stile sono stati realizzati da Fabio Ciotti. Per una descrizione dettagliata di TEI/XML si vedano il sito della TEI (<http://www.tei-c.org/>) dove sono disponibili le *Guidelines: TEI P5* (2015). Per un'introduzione più agile si veda Burnard, Sperberg-McQueen 2005, la cui edizione italiana è curata da Fabio Ciotti.

21 Si è fornita anche la possibilità di scaricare l'intero corpus in formato TEI/XML.

22 <hi rend="italic"> per il corsivo, <hi rend="bold"> per il grassetto e così via.

La marcatura TEI/XML è gerarchica e non permette la coesistenza di due strutture concorrenti di pari livello. Tuttavia il problema viene in parte superato con l'uso di marcatori vuoti che, invece di incorniciare la porzione testuale da descrivere, si limitano a essere situati nel punto di inizio di un determinato fenomeno.

I capitoli, ad esempio, sono marcati come `<div type="cap" n="">` all'inizio del capitolo e `</div>` alla fine. Se il capitolo contiene paragrafi, questi sono marcati come una `<div>` di livello inferiore `<div type="par" n="">` contenuto del paragrafo `</div>`. La marcatura di capitoli e paragrafi come `<div>` è molto comoda perché consente di scandire il testo in una struttura precisa, prevede per ogni `<div>` la possibilità di inserire titoli (`<head>`) e permette la creazione di alberi di navigazione funzionali alla lettura a schermo. Nei casi in cui, tuttavia, le numerazioni di capitoli e paragrafi non sono gerarchizzate, ma indipendenti e intrecciate, la marcatura con le `<div>` diventa impossibile:

```
<div type="cap" n="1">
<div type="par" n="1">testo del primo paragrafo</div>
<div type="par" n="2">testo del secondo paragrafo</div>
<div type="par" n="3"><p> inizio del testo del terzo paragrafo
</div>
<div type="cap" n="2">
proseguimento del testo del terzo paragrafo</p></div>
<div type="par" n="4"><p>testo del quarto paragrafo </p></div>
</div>
```

Per ovviare a tale problema, come si è detto, sono previsti marcatori puntuali. Ad esempio la biblioteca *digilibLT*, nei casi in cui si presenta conflitto, marca i paragrafi come `<milestone unit="par" n=""/>`. E, come si è accennato, il conflitto si genera non solo quando capitoli e paragrafi si incrociano, ma anche quando essi non coincidono con i capoversi.

La marcatura in `<div>` prevede, infatti, che il testo sia inserito non direttamente fra il tag di apertura e di chiusura delle `<div>` relative, ma all'interno del tag che delimita il capoverso `<p>` e che implica a capo (`<div><p> testo </p></div>`); ne consegue che il cambio di `<div>` determina l'inserimento di un capoverso. Il tag `<milestone/>`, invece, non implica l'inserimento di un capoverso, ma rispetto alle `<div>` è più limitante sia per l'approfondimento della marcatura sia per la creazione di alberi di navigazione sia per la possibilità di isolare con precisione la porzione testuale a cui fa riferimento.

Se nel caso di conflitto tra capitoli e paragrafi la scelta del `<milestone/>` al posto della `<div>` è obbligata, a meno di rinunciare a marcare una delle due numerazioni, nel caso in cui vi sia solo mancata coincidenza tra paragrafi/capitoli) e capoversi, vien da domandarsi se valga la pena di

rinunciare alle potenzialità della marcatura con <div> per non inserire degli a capo in punti non previsti dall'edizione di riferimento e spesso non in coincidenza con segni di interpunzione forte.

In realtà non esiste una risposta valida in assoluto; molto dipende, infatti, dalla finalità per cui si sono marcati i testi e dal prodotto finale che si vuole ottenere. Se i testi confluiscono semplicemente in una banca dati interrogabile, il fatto che vi sia una corretta scansione in capoversi diventa di secondaria importanza. Se invece si prevede di ricavare dall'XML anche formati in cui l'impaginazione ha un ruolo fondamentale, perché sono pensati per la lettura e la stampa e non solo per essere interrogati dalla macchina, allora la prospettiva è diversa. La scelta di voler ricavare PDF e ePub dall'XML ha implicato una scelta di marcatori che non ponesse limitazioni alle esigenze di visualizzazione. In tali formati, infatti, sarebbe fuori luogo un testo che andasse a capo a metà frase solo perché in quel punto vi è un cambio di paragrafo.

Nel .txt, invece, sarebbe irrilevante sia perché il formato non ha come utilizzo primario la lettura sia perché, essendo un formato editabile, può essere agevolmente modificato. Anch'esso, tuttavia, non è esente da problemi connessi alla visualizzazione, per quanto essi siano di tutt'altro genere.

Come è noto questo formato non consente di conservare le formattazioni, ma tale mancanza avrebbe comportato una perdita di significati non trascurabile. Le edizioni di riferimento dei testi della biblioteca *digilibLT* sono eterogenee e usano formattazioni e diacritici con i più svariati significati. Abbiamo provveduto a normalizzare i diacritici più comuni e le formattazioni di più semplice interpretazione,²³ ma non è stato possibile uniformare completamente tutti i testi perché sarebbe stato necessario uno studio approfondito per ognuno di essi, e ciò avrebbe allungato di molto i tempi di pubblicazione.²⁴ Si è quindi deciso di utilizzare i marcatori

23 Ad esempio l'uso del grassetto o del corsivo per indicare i titoli di capitolo o paragrafo. In questi casi si è adottata la marcatura <head> e attraverso i fogli di stile i titoli si visualizzano in grassetto. Una sintesi delle normalizzazioni effettuate è contenuta nella nota al testo inclusa in TEI Header nell'<editorialDecl> e leggibile nei frontespizi dei formati PDF e ePub, nonché nella pagina online dedicata alle informazioni editoriali di ogni testo. Chi desiderasse maggiori dettagli può inoltre consultare le *Note di trascrizione e di codifica* (scaricabili dal sito all'indirizzo http://www.digiliblt.uniupo.it/opere_tardoantico/Note_di_trascrizione_e_di_codifica.pdf) e Musso 2014, 172-5.

24 Una distinzione di tutti gli usi del corsivo (integrazione, integrazione solo in presenza di lacuna materiale, correzione del testo dei codici, citazione, titolo di opera, nomi di autori, ecc.) o delle altre tipologie di formattazione, con relativa marcatura specifica, avrebbe rallentato il lavoro senza garantire una soluzione ottimale a ogni esigenza e avrebbe impedito di conseguire lo scopo principale del progetto: mettere a disposizione degli studiosi, in tempi brevi, un vasto corpus di testi liberamente ricercabile e scaricabile. Approfondimenti di marcatura sono sempre possibili in una seconda fase. Vengono comunque specificati per ogni testo eventuali usi particolari della formattazione.

previsti dalla TEI per descrivere le formattazioni: `<hi rend="">`; il valore dell'attributo `@rend` indica il tipo di formattazione (italic, bold, ecc.).

Ma anche qualora avessimo attribuito una marcatura descrittiva specifica a ogni fenomeno, in fase di visualizzazione si sarebbe comunque dovuto ricorrere alle formattazioni per indicare che quella porzione ha un valore peculiare. Ne consegue, quindi, che ogni informazione rappresentata dalla formattazione sarebbe andata persa nel .txt.

Per evitare tale impoverimento, si è deciso di utilizzare una sequenza di caratteri speciali come indicatori delle diverse formattazioni (si veda la tabella 1).

Formattazione	Marcatura	Segnalazione in .txt
Corpo minore	<code><hi rend="small">testo</hi></code>	<code>{@testo@}</code>
Corsivo	<code><hi rend="italic">testo</hi></code>	<code>{€testo€}</code>
Grassetto	<code><hi rend="bold">testo</hi></code>	<code>{\$testo\$}</code>
Sottolineato	<code><hi rend="underline">testo</hi></code>	<code>{#testo#}</code>
E s p a n s o	<code><hi rend="expanded">testo</hi></code>	<code>{%testo%}</code>

Tabella 1. Indicatori delle diverse formattazioni

Tale prospetto si trova in testa ai file .txt, affinché l'utente sia informato di ogni modifica apportata ai testi presenti nella biblioteca.

4 Glifi

Riflessioni sulle modalità di visualizzazione sono state necessarie anche per rappresentare al meglio quei simboli non ancora presenti nello standard Unicode o, se presenti, non inclusi nei font più comuni. Per quanto molto ricco, Unicode inevitabilmente non può prevedere e comprendere ogni glifo esistente. Tali simboli sono particolarmente numerosi nei testi in prosa tardoantica, data la natura tecnica di molti di essi, e per lo più si possono raggruppare in due tipologie: numerali con moltiplicatori e unità di misura, per i quali si sono adottate due soluzioni distinte.²⁵

Per i numerali si è scelto di esplicitarne il valore numerico. A testo si scrive il numerale senza moltiplicatore e con la marcatura si dichiara

²⁵ Inizialmente si era pensato di adottare un font gratuito creato appositamente per studiosi di testi antichi (Cardo) e di segnalare agli utenti della biblioteca *digilibLT* il sito da cui è possibile scaricarlo. Tale opzione, però, ha presto mostrato inconvenienti. Anzitutto neppure Cardo è in grado di soddisfare ogni esigenza, dato che non rappresenta tutti i simboli necessari. Inoltre, i glifi non presenti in Unicode sono stati inseriti nell'area ad uso privato comportando un'errata visualizzazione non solo a chi non ha scaricato la versione più aggiornata di Cardo, ma anche nel formato TXT ricavato dall'XML. Si è quindi dovuta cercare una soluzione alternativa utilizzando la marcatura prevista dalla TEI.

come deve essere interpretato. Per rappresentare, ad esempio, il simbolo indicante 10000, X con linea orizzontale soprascritta, si scrive:

```
<num value="10000">X</num>
```

la cui visualizzazione è:

X(10000)

Nel caso di unità di misura, e in generale per i simboli che non rappresentano numeri, è necessaria una marcatura più articolata. In TEI Header si descrivono i vari caratteri fornendo a ciascuno un identificatore univoco.

```
<encodingDesc>
<charDecl>
<char xml:id="char1">
<charName>Nome del simbolo</charName>
<desc>Descrizione dell'aspetto del simbolo</desc>
</char>
<char xml:id="char2">
<charName>Nome del simbolo</charName>
<desc>Descrizione dell'aspetto del simbolo</desc>
</char>
</charDecl>
</encodingDesc>
```

A testo si rinvia al carattere in questione col tag vuoto <g/>. Qualora il simbolo esista, anche se solo nell'area a uso privato di un font specialistico, esso viene inserito all'interno dell'elemento <g>:

```
<text>
<body>
<div>
<p>testo <g ref="#char1">—</g> testo. </p>
<p>testo <g ref="#char2"/> testo. </p>
</div>
</body>
</text>
```

Attraverso i fogli di stile si è stabilito che accanto al simbolo, o al suo posto, comparisse tra parentesi il nome dell'unità di misura, per ognuna delle quali si è stabilita un'abbreviazione, il cui scioglimento è riportato nel marcatore <editorialDecl>.

Den. per Denarius
 Drag. per Dragma
 Pon. per Pondus
 Scrip. per Scripulus
 Sem. per Semis
 Semunc. per Semuncia
 Sext. per Sextarius
 Vict. per Victoriatus
 Vnc. per Vncia
 ecc.

Vediamo un esempio sia di unità di misura per cui Unicode (o un font nell'area a uso privato) prevede il simbolo sia di unità per cui il simbolo è assente.

In TEI Header, nel <charDecl>, si esplicitano gli identificatori dei simboli, per ognuno dei quali si aggiungono l'abbreviazione del nome e la descrizione del suo aspetto.

```
<charDecl>
<char xml:id="char1">
<charName>Vnc.</charName>
<desc>Tractus longus</desc>
</char>
<char xml:id="char2">
<charName>Sext.</charName>
<desc>S cum linea transuersa</desc>
</char>
</charDecl>
```

A testo, nel marcatore <g>, si richiama l'identificativo del glifo in questione:

```
<milestone unit="par" n="30"/> Anethi surculi uruntur eiusque pulueris cribrati <g ref="#char1">—</g> I, cicharbae <g ref="#char1">—</g> III remittes in olei optimi <g ref="#char2"/> I et adicies gallae tunsae semunciam inlinesque eo porriginosum et ulcerosum capite contra pilum et postea lauabis in balneo.
```

Il testo riportato ha la seguente visualizzazione:

[30] Anethi surculi uruntur eiusque pulueris cribrati —[Vnc.] I, cicharbae —[Vnc.] III remittes in olei optimi [Sext.] I et adicies gallae tunsae semunciam inlinesque eo porriginosum et ulcerosum capite contra pilum et postea lauabis in balneo.

Tale marcatura da un lato supplisce a una carenza di Unicode e dall'altra permette di attuare una schedatura dei simboli presenti nelle edizioni,²⁶ operazione necessaria per poter chiedere a Unicode di ampliare lo standard.

5 Natura collaborativa e qualità: un rapporto problematico

Le questioni prese in esame mostrano le scelte adottate dalla biblioteca *digilibLT* di fronte alle problematiche sollevate dal processo di digitalizzazione. Va da sé che si tratta di una delle varie possibilità offerte dalla TEI, che non impone uno schema rigido e univoco di marcatura. Anzi, la sua natura aperta permette anche di introdurre nuovi marcatori purché siano dichiarati attraverso DTD (*Document Type Definition*) o XSD (*XML Schema*). Se tale versatilità permette di adattare al meglio la marcatura alle esigenze di chi digitalizza, è pur vero che non favorisce la compatibilità fra testi realizzati in progetti diversi, che pur utilizzano il vocabolario TEI. Non mancano tentativi di creare vocabolari condivisi per testi di tipologie specifiche, come dimostra *Epidoc* per i documenti epigrafici antichi,²⁷ tuttavia si è ancora lontani dalla possibilità di creare un unico database testuale interrogabile nel suo complesso partendo dai testi latini presenti nelle varie biblioteche digitali liberamente consultabili online.

Questo non significa, tuttavia, che non vi possano essere proficui rapporti di collaborazione fra progetti diversi o fra studiosi.²⁸ La biblioteca *digilibLT* ha infatti creato una rete di contatti con specialisti del settore, non solo italiani, che hanno fornito il proprio contributo a vari livelli.

È necessario chiarire le modalità in cui avviene il rapporto di collaborazione. Il web odierno tende sempre più a favorire l'inserimento diretto di dati da parte dell'utente, ma il gruppo di ricerca ha preferito esercitare un controllo accurato delle informazioni presenti nella biblioteca prima della loro immissione per garantire non solo, ovviamente, la correttezza dei dati, ma anche la coerenza e l'uniformità nelle modalità di preparazione dei dati.

Abbiamo già menzionato l'importanza delle segnalazioni ricevute nella compilazione del canone (§ 3) e l'arricchimento per la biblioteca prodotto da schede realizzate dai massimi specialisti del testo in questione (§ 2.3).

26 Tale schedatura è stata avviata da David Paniagua, esperto di testi tecnici e progressivamente ampliata dai collaboratori della biblioteca *digilibLT* via via che si sono riscontrati nuovi simboli e glifi. La complessità di tale schedatura è accentuata dal fatto che le varie edizioni critiche utilizzano glifi diversi per rappresentare la stessa unità di misura. Riportiamo a titolo esemplificativo, tra parentesi quadre, alcune rappresentazioni dell'oncia [— · / (~)].

27 Le *Guidelines* sono reperibili all'indirizzo <http://www.stoa.org/epidoc/gl/latest/>.

28 Il sito della biblioteca prevede una pagina specifica di contatti e feedback (<http://www.digiliblt.uniupo.it/contatti.php>).

In alcuni casi, inoltre, il testo digitalizzato è stato preparato da chi ne sta curando l'edizione critica, facendo sì che l'edizione digitale presenti un testo criticamente più accurato delle edizioni cartacee esistenti.²⁹

Infine sono preziosi i contributi e le segnalazioni ricevuti per eliminare refusi rimasti nella preparazione dei testi. Garantire la massima correttezza dei testi è stato un obiettivo centrale del progetto. A tal scopo sono state valutate le varie procedure di digitalizzazione per adottare quella che garantisse la minor incidenza di errore. La biblioteca *digilibLT* ha consapevolmente scelto l'opzione controtendenza di mantenere all'interno del gruppo di ricerca tutte le fasi di digitalizzazione, compreso il controllo del testo da marcare, non demandando questa fase a ditte esterne. Si è già avuto modo di riflettere su tale questione,³⁰ in questa sede ci soffermeremo solo su alcuni aspetti. Si tenga presente, come si è detto, che la biblioteca è partita da zero, senza un corpus già digitalizzato (a eccezione dei grammatici). Gran parte del tempo e delle energie è quindi stata dedicata a tale operazione preliminare.

Per creare un testo digitale sostanzialmente vi sono due vie: o viene riscritto, o viene resa editabile l'immagine dell'edizione cartacea del testo dell'edizione di riferimento.³¹ La prima strada non era sostenibile per un gruppo di dimensioni ridotte, si è quindi optato per la seconda via, puntando a ottenere un'alta qualità delle immagini digitalizzate attraverso l'uso di scanner professionali *Atiz BookDrive*. Tali scanner sono caratterizzati da un piano a V su cui poggia il libro aperto e ognuno dei due piatti viene fotografato attraverso macchine reflex full frame (nello specifico si tratta di Canon 5D) che permettono di ottenere immagini ad altissima risoluzione. I software in dotazione (*BookDrive Capture* e *BookDrive Editor*) consentono non solo di regolare le impostazioni delle macchine in fase di scatto, ma anche di ottimizzare le immagini ottenute per migliorare la resa dell'OCR.³²

Nonostante questo, l'incidenza di errore dei programmi di riconoscimento ottico dei caratteri è rimasta elevata. Ciò si spiega non solo con i difetti meccanici della fonte cartacea, ma anche con le difficoltà linguistiche presentate dai testi. Anzitutto per il latino tardo non ci sono validi dizionari di supporto al programma di riconoscimento, senza contare che, laddove un errore generasse una parola esistente, nessun dizionario potrebbe se-

29 Si veda, ad esempio, il caso del commento di Servio all'*Eneide* di Virgilio, che *DigilibLT* (per gentile concessione dell'editore stesso) offre in anteprima nell'edizione di Giuseppe Ramires.

30 Cf. Musso 2014, 154-7 e per una riflessione più generale Lana 2012a, 76-83.

31 Togliendo ovviamente tutto ciò che non è testo antico: apparato, note, introduzione, commento, ecc.

32 Ad esempio aumentando il contrasto laddove vi siano pagine sbiadite o ingiallite.

gnalarlo. Le grafie non canoniche, l'uso del greco e di numerosi simboli, inoltre, contribuiscono a ostacolare un buon riconoscimento dei caratteri.

È emersa immediatamente la necessità di un approfondito controllo umano all'esito dell'OCR. Inizialmente si pensava di poter evitare una lettura integrale dei testi procedendo al confronto automatico di due distinti riconoscimenti della stessa opera.³³ Tale espediente funziona bene se il numero di errori è limitato ma, nel nostro caso, l'alta incidenza d'errore avrebbe costretto alla revisione di quasi tutte le righe di testo. Si è quindi optato per una duplice lettura integrale dei testi, affidata a studiosi e studenti che avessero una buona competenza in lingua latina. Quest'ultimo particolare ci ha permesso in numerosi casi di segnalare, ad esempio, errori tipografici presenti nell'edizione di riferimento.

Ovviamente siamo ben consci del fatto che nemmeno una duplice lettura può garantire l'individuazione di tutti i refusi. Ed è per questo motivo che assume grande importanza il controllo ulteriore che possono esercitare gli utenti sui testi online. Si sono rivelate di grande utilità le segnalazioni effettuate da un collaboratore del *Thesaurus linguae Latinae*, Johann Ramminger, che attraverso un'analisi automatica dei testi ha isolato una serie di forme sospette. In alcuni casi si trattava effettivamente di refusi o di errori già presenti nell'edizione di riferimento ma molte, in realtà, erano forme tarde e ciò ha messo in chiara evidenza il valore rappresentato dal corpus della biblioteca *digilibLT* per uno studio dell'evoluzione della lingua latina.

6 Digitale, perché?

Se, ad esempio, nel caso delle dispense universitarie degli anni '60, stampate su carta acida, il mezzo digitale rappresenta uno strumento di salvaguardia della memoria, abbiamo però spesso notato come la rapidissima evoluzione della tecnologia esponga i prodotti digitali della ricerca all'elevato rischio dell'obsolescenza dei supporti, che non sempre riescono ad adattarsi alle diverse versioni dei sistemi operativi. Vi sarebbe infatti bisogno di una manutenzione costante e a lungo termine, pratica virtuosa ma resa poco probabile dalla scadenza dei progetti o dalla scarsità di fondi. Nasce così un acceso dibattito all'interno del mondo accademico, che se da un lato riconosce il valore dello strumento digitale in termini di velocità di accesso e come vettore di rapida diffusione della cultura, non di rado guarda ancora alla *digital scholarship* con qualche perplessità,³⁴ relativa

33 Attraverso programmi specifici quali ad esempio WinMerge.

34 In merito, ancora valide le riflessioni di McGann 2005, 77, secondo cui i progetti digitali «typically get born into poverty - even the best-funded ones. Ensuring their maintenance,

non solo al problema della caducità dei supporti, ma anche a quello – assai più controverso – del controllo scientifico e del valore di quanto prodotto con metodologie digitali in varie sedi (cf. Tabacco 2014, 7-8; Musso 2014, 186-7).

La progettazione ex novo di *DigilibLT*, che non aveva alle spalle una banca dati in CD-ROM, neppure in forma di prototipo, ha altresì favorito una riflessione complessiva non solo in termini tecnici, ma anche ma anche sulle potenzialità del progetto ad ampio raggio. Si è così scelto, come già è stato detto, di mantenere all'interno del progetto tutti i passaggi della lavorazione dei testi, una decisione che ha quindi imposto l'allargamento del gruppo di ricerca. A questo fine *DigilibLT* ha guardato in modo particolare ai giovani, facendone tuttavia, non una mera forza lavoro, ma costruendo intorno a loro un vero e proprio percorso formativo in filologia digitale.³⁵

6.1 Un punto di partenza

Una banca dati online marcata secondo XML-TEI non rappresenta, tuttavia, solo un risultato raggiunto: essa deve anche costituire un punto di partenza per nuovi approcci e ricerche. Già nel 2003 John Unsworth tracciando un bilancio dello sviluppo fino ad allora delle *digital humanities* affermava:

we need (we still need) to demonstrate the usefulness of all the stuff we have digitized over the last decade and more – and usefulness not just in the form of increased access, but specifically, in what we can do with the stuff once we get it: what new questions we could ask, what old ones we could answer.

Più di dieci anni dopo, queste domande rimangono attuali:

a tuttora, l'enfasi sembra essere stata posta prevalentemente sulla produzione, piuttosto che sull'impiego dei dati così prodotti, come se si trattasse soltanto di formare un immenso archivio di testi codificati,

development, and survival is a daunting challenge. Worst of all, the work regularly passes without much practical institutional notice. Accepted professional standards do not control the work in objective ways. Most of it comes into being without oversight or peer review». Si veda anche la lettera inviata dal direttivo dell'AIUCD (Associazione per l'Informatica Umanistica e la Cultura Digitale), Dino Buzzetti nel febbraio 2014 all'allora ministro dell'Istruzione, Università e Ricerca Stefania Giannini in merito alla questione della valorizzazione, anche in sedi concorsuali, di quanto prodotto con metodologie digitali (per il testo della lettera cf. <http://www.aiucd.it/osservazioni-critiche-dellaiucd-sullasn/>).

³⁵ Del valore formativo di *DigilibLT* si è già dato conto in Borgna 2017, 13-16.

predisposti per così dire a ‘futura memoria’ senza considerare più di tanto il problema dell’uso e delle modalità di impiego di tali risorse. (Buzzetti 2011, 39)

Il sostegno di *DigilibLT* all’*open access* come canale di disseminazione della ricerca, una scelta obbligata laddove tale ricerca sia sostenuta da finanziamenti pubblici, è così in grado di dare origine ad altre iniziative, nella volontà di rifuggire quell’isolamento a cui è spesso soggetta la *Digital Scholarship*.³⁶ In questo senso la biblioteca si è posta come base di partenza per il progetto TBL, *Textual Bilingualism in Latin* che indaga i fenomeni di contatto linguistico tra greco e latino nei testi letterari latini tardi.³⁷ L’esperienza maturata e la strumentazione acquisita ha poi portato alla fondazione, presso l’Università di Torino (proponente aggiuntivo di *DigilibLT*), del Centro di Ricerca Interdipartimentale per la digitalizzazione e la realizzazione di Biblioteche Digitali Umanistiche MEDIHUM (*Memoria Digitalis Humanistica*).³⁸ Da ricordare anche GeoLat, Geografia per la letteratura latina, che in vista di un accesso ai testi della latinità tarda a partire da mappe, ha elaborato e recentemente diffuso Go!, la prima ontologia geografica esplicitamente studiata per il mondo antico.³⁹

In conclusione, alla domanda sempre più pressante di istituire un circolo virtuoso tra risorse letterarie online, ontologie e *linked data*, come grande opportunità per lo sviluppo futuro non solo delle *Digital Humanities*, ma degli studi umanistici in genere (cf. Ciotti 2014, 9), *DigilibLT* sta quindi rispondendo non solo con la sua stessa costante crescita, ma anche ponendosi come motore di sviluppo per altre iniziative, segno che l’investimento sui giovani e l’attenzione per l’aspetto collaborativo e l’accesso aperto (mai privi, tuttavia, di un rigoroso controllo scientifico) si stanno rivelando intuizioni corrette.

36 «Digital scholarship - even the best of it - is all more or less atomized, growing like so many Topsyies. Worse, these creatures are idiosyncratically designed and so can’t easily talk to each other» (McGann 2005, 77).

37 Cf. <http://tbl.uniupo.it> (direzione scientifica: Maria Napoli)

38 Cf. <http://www.medihum.unito.it>. Il centro coniuga esigenze di usabilità e accesso comune ai contenuti culturali e di ricerca, digitalizzati o nativi digitali, con l’orizzonte scientifico delle attività di text e data mining e di linguistica computazionale per l’estrazione e l’analisi dei dati testuali

39 Cf. <https://goo.gl/3VRPGt> (direzione scientifica: Maurizio Lana). In merito si vedano anche Ciotti et al. 2016.

Bibliografia

- Borgna, Alice (2016): «Latino e informatica. Una proposta di et... et nella società dell'aut... aut». *ClassicoContemporaneo*, 3 (2017), 1-22.
- Burnard, Lou; Sperberg-McQueen, Michael (2005). *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari* [online]. A cura di Fabio Ciotti. Milano: Sylvestre Bonnard. URL http://www.tei-c.org/Vault/P4/Lite/teiu5_it.html (2017-10-17).
- Buzzetti, Dino (2011). «Oltre il rappresentare. Le potenzialità del *markup*». Fiermonte, Perilli 2011, 39-62.
- Ciotti, Fabio (2014). «Digital Literary and Cultural Studies. State of the Art and Perspectives». *Between*, 8(4), 1-17.
- Ciotti, Fabio et al. (2016). «Ontologies and the Cultural Heritage. The case of GO!» [online]. Draelants, Isabelle et. al. (eds.), *Proceedings of the II International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016)*, 7-18. URL <http://ceur-ws.org/Vol-1595/> (2017-11-10).
- Eadie, John William (1967). *The Breviarium of Festus. A Critical Edition with Historical Commentary*. London: Athlone Press.
- Fele, Maria Luisa (2009). *Il breviarium di Rufio Festo*. Hildesheim: Weidmann.
- Fioromonte, Domenico; Perilli, Lorenzo (a cura di) (2011). *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*. Firenze: Le Lettere.
- Guidelines: TEI P5 (2015). «Guidelines for Electronic Text Encoding and Interchange» [online]. URL <http://www.tei-c.org/Guidelines/> (2017-10-17).
- Lana, Maurizio (2011). «Un *database* testuale per il latino tardo». Fioromonte, Perilli 2011, 281-300.
- Lana, Maurizio (2012a). *Biblioteche digitali. Un'introduzione*. Bologna: Bononia University Press.
- Lana, Maurizio (2012b). «Da una *digital library* del latino tardo ad un corpus globale». Ciotti, Fabio; Crupi, Gianfranco (a cura di), *Dall'informatica umanistica alle culture digitali = Atti del convegno di studi in memoria di Giuseppe Gigliozzi* (Roma, 27-28 ottobre 2011). Roma: Digilab, 134-50.
- Lana, Maurizio (2012c). «Metodologie e problematiche per una biblioteca digitale. Il caso di *digilibLT*». *DigItalia. Rivista del digitale nei beni culturali*, 7, 40-64.
- Lomanto, Valeria (1980). «Lessici latini e lessicografia automatica». *Memorie dell'Accademia delle Scienze di Torino. Classe di Scienze Morali, Storiche e Filologiche*, V(4), 113-270.
- Marinone, Nino (1979). «Concordanza dei Grammatici Latini». Suppl. di *Atti dell'Accademia delle Scienze di Torino. Classe di Scienze Morali, Storiche e Filologiche*, 112, 9-19.

- McGann, John J. (2005). «Culture and Technology. The Way We Live Now, What Is to Be Done?». *New Literary History*, 36(1), 71-82.
- Mosca, Laura; Musso, Simona (2016). «Costruire una biblioteca digitale di testi latini tardoantichi. Problemi e scelte». Maconi, Ludovica; Marazzini, Claudio (a cura di), *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori = Atti del Convegno "Piazza delle Lingue", VIII Edizione* (Firenze 6-7 novembre 2014). Firenze: Edizioni dell'Accademia della Crusca, 143-58.
- Musso, Simona (2014). «La marcatura di testi latini tardoantichi. Un compromesso ragionevole». *Aevum Antiquum*, 11, 151-87.
- Tabacco, Raffaella (2000). *Itinerarium Alexandri*. Testo, apparato critico, introduzione, traduzione e commento a cura di Raffaella Tabacco. Firenze: Leo S. Olschki Editore.
- Tabacco, Raffaella (2014). «La tarda antichità latina tra i codici e il web». *Aevum Antiquum*, 11, 3-17.
- Tabacco, Raffaella (2016). «La prosa latina pagana tardoantica e la biblioteca digitale *digilibLT*». *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori = Atti del Convegno "Piazza delle Lingue", VIII Edizione* (Firenze 6-7 novembre 2014). Firenze: Edizioni dell'Accademia della Crusca, 125-41.
- Unsworth, John (2003). «Tool-Time, or 'Haven't We Been Here Before?'. Ten Year in Humanities Computing» [online]. *Transforming Disciplines. The Humanities and Computer Science* (18th January 2003). Washington DC. URL <http://people.virginia.edu/~jmu2m/carnegie-ninch.03.html> (2017-11-10).