

Antichistica 14
Filologia e letteratura 3

Strumenti digitali e collaborativi per le Scienze dell'Antichità

a cura di
Paolo Mastandrea



Edizioni
Ca' Foscari

Risorse digitali per le Scienze dell'Antichità

Antichistica
Filologia e letteratura

Collana diretta da
Lucio Milano

14 | 3



Edizioni
Ca' Foscari

Antichistica

Filologia e letteratura

Direttore scientifico

Lucio Milano (Università Ca' Foscari Venezia, Italia)

Comitato scientifico

Claudia Antonetti (Università Ca' Foscari Venezia, Italia)

Filippo Maria Carinci (Università Ca' Foscari Venezia, Italia)

Ettore Cingano (Università Ca' Foscari Venezia, Italia)

Joy Connolly (New York University, USA)

Andrea Giardina (Scuola Normale Superiore, Pisa, Italia)

Marc van de Mieroop (Columbia University in the City of New York, USA)

Elena Rova (Università Ca' Foscari Venezia, Italia)

Fausto Zevi (Sapienza Università di Roma, Italia)

Direzione e redazione

Dipartimento di Studi Umanistici

Università Ca' Foscari Venezia

Palazzo Malcanton Marcorà

Dorsoduro 3484/D,

30123 Venezia

Strumenti digitali e collaborativi per le Scienze dell'Antichità
Paolo Mastandrea (a cura di)

© 2017 Paolo Mastandrea per il testo

© 2017 Edizioni Ca' Foscari - Digital Publishing for the present edition



Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale
This work is licensed under a Creative Commons Attribution 4.0 International License

Qualunque parte di questa pubblicazione può essere riprodotta, memorizzata in un sistema di recupero dati o trasmessa in qualsiasi forma o con qualsiasi mezzo, elettronico o meccanico, senza autorizzazione, a condizione che se ne citi la fonte.

Any part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without permission provided that the source is fully credited.

Edizioni Ca' Foscari - Digital Publishing
Università Ca' Foscari Venezia
Dorsoduro 3246, 30123 Venezia
<http://edizionicafoscari.unive.it/> | ecf@unive.it

1a edizione novembre 2017
ISBN 978-88-6969-182-9 [ebook]
ISBN 978-88-6969-183-6 [print]

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari - Digital Publishing: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della collana. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.

Scientific certification of the works published by Edizioni Ca' Foscari - Digital Publishing: all essays published in this volume have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Scientific Committee of the series. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

URL <http://edizionicafoscari.unive.it/it/edizioni/libri/978-88-6969-183-6/>
DOI 10.14277/978-88-6969-182-9

Oltre le biblioteche digitali

Maurizio Lana

(Università degli Studi del Piemonte Orientale, Italia)

Abstract Geolat (Geography for Latin Literature) is a research project in its startup phase aimed at building a collaborative scholarly reading of classical Latin texts enhancing the reuse of their informative content. The core concept is that the texts are annotated and that the annotation in turn can be used to study the texts. The TEI conforming annotation is aimed to place-names and employs an ad hoc geographical ontology for the world-view of Classical texts. The re-use of the annotation can start from a map interface mixed with a faceted textual search allowing to ask new questions, e.g. which is the geography of Horace's *Odes* and what does it mean? Has it any relation with the geography of other Augustan age authors? Has it any functional or ideological meaning? This basic concept can be expanded adding more ontologies for person names/roles and events/times; or also extending it to more literatures of the European area for an in-depth study of places, roles and events across times.

Sommario 1 La digitalizzazione del patrimonio culturale e la creazione di biblioteche digitali. – 1.1 Segno dell'avanzata pervasiva del digitale. – 1.2 Costosa nell'avvio. – 1.3 Costosa nella manutenzione. – 2 A che cosa serve una biblioteca digitale? – 3 Attività consuete, attività nuove. – 4 Licklider e Engelbart: *Augmenting Human Intellect*. – 5 Dalla ricerca di parole alla ricerca di categorie: l'annotazione semantica. – 6 Un caso concreto: Geolat e la biblioteca digitale digilibLT. – 6.1 digilibLT. – 6.2 Geolat. – 7 Per concludere.

Keywords Digital libraries. Geography. Semantic annotation. Ontologies. Named entities recognition.

1 La digitalizzazione del patrimonio culturale e la creazione di biblioteche digitali

La spinta alla digitalizzazione del patrimonio culturale è continua e pressante. Il patrimonio culturale è delicato e la sua natura fisica ne limita la fruizione a chi si trovi in sua presenza: invece la digitalizzazione prospetta la possibilità che, da un lato, il pubblico fruisca di copie degli originali che rimangono intatti e non soggetti ad usura, dall'altro che molteplici fruitori dispersi nel mondo possano accedere al medesimo bene nella sua forma digitalizzata. Le biblioteche digitali sono un caso paradigmatico di questa tendenza.

1.1 Segno dell'avanzata pervasiva del digitale

Il contesto in cui si colloca la creazione di biblioteche digitali è quello del mondo digitale. Negli anni '90 del secolo scorso l'accesso remoto ai cataloghi delle biblioteche tramite strumenti software di difficile uso (Telnet) appariva un'opportunità incredibilmente avanzata e capace di incidere significativamente nelle pratiche della ricerca in quanto eliminava la necessità di spostarsi fisicamente da una città all'altra per consultare i cataloghi cartacei. Oggi, per chiunque davanti all'esito di una ricerca all'interno di un catalogo bibliotecario c'è l'istinto di cliccare sul titolo per vedere aprirsi in lettura il testo dell'opera - poi la consapevolezza razionale interviene a dire che non sempre questo è possibile. Ma il web ha insegnato questo: che le risorse, i contenuti, i testi (le musiche, i film, le foto) sono a portata di un clic del mouse. La digitalizzazione delle biblioteche rientra in questo grande quadro complessivo di presenza del web nelle vite delle persone e degli studiosi, o per meglio dire di intersezione sempre più inestricabile di mondo fisico e mondo digitale. Poiché parti sempre più ampie e importanti delle vite delle persone si svolgono nel mondo digitale, si tende a portare sempre più nel mondo digitale strumenti e risorse nate nel (per il) mondo fisico.

1.2 Costosa nell'avvio

Restringendo l'attenzione dal tema ampio della digitalizzazione del patrimonio culturale a quello più specifico delle biblioteche e delle biblioteche digitali, occorre in primo luogo notare che i costi da sostenere per digitalizzare il contenuto di una biblioteca sono molto alti, pertanto quasi mai i fondi presenti nel bilancio ordinario sono sufficienti allo scopo e servono quindi finanziamenti ad hoc. Per avere una percezione dell'entità delle cifre in gioco basti pensare al fatto che la digitalizzazione di una parte del posseduto della Biblioteca Nazionale Centrale di Firenze è stata intrapresa da Google per arricchire la biblioteca digitale Google Books - e si sta parlando di una delle più grandi aziende operanti nel mondo digitale (mentre lo Stato italiano non è in condizione di progettare e realizzare in proprio un'iniziativa di questo genere per la conservazione e la diffusione del proprio patrimonio librario). Iniziative più ridotte, mirate ad ambiti disciplinari specifici, sono comunque possibili anche da parte di soggetti più piccoli: per restare in ambito italiano le ben note biblioteche digitali di contenuto latinistico/classicistico Muisque Deoque, digilibLT, BIA, ALIM, CoRiMu, e altre non hanno avuto bisogno per nascere della presenza di grandi aziende del mondo digitale; ma tutte si sono basate (o si basano) su finanziamenti di ricerca specifici.

1.3 Costosa nella manutenzione

Le biblioteche digitali non solo presentano alti costi di avvio ma sono anche molto più delicate e complesse delle biblioteche fisiche. Questa delicatezza si può esemplificare in modo evidente pensando alla sopravvivenza nel tempo: possediamo oggi biblioteche create secoli e secoli fa e siamo in grado di leggere agevolmente le opere che esse contengono; mentre non riusciamo più a leggere il file di un articolo scritto 30 anni fa con WordStar. Si pensi dunque a quanto più delicato sia il funzionamento di una biblioteca digitale che per operare, cioè per mettere le opere a disposizione dei lettori, ha bisogno di un server, cioè di un computer dotato di sistema operativo e programmi, connesso in rete; e ognuno degli elementi dell'insieme (computer, sistema operativo, programmi, connessione in rete) essendo esposto a rapida obsolescenza richiede un costante lavoro di manutenzione e aggiornamento. Ma le parti sono interconnesse e quindi può accadere che l'aggiornamento di una di esse implichi la necessità di aggiornarne altre (e.g. un programma che non funziona più con una nuova versione del sistema operativo) che potrebbero però non essere (più) disponibili in versione compatibile. Poi ci sono i formati dei dati, che evolvono nel tempo e sono anch'essi soggetti a obsolescenza; e i formati di scrittura dei supporti; e gli standard fisici, elettrici, digitali, delle periferiche che leggono i supporti (i connettori degli hard disk ne sono un esempio).

Non essendo praticabile né efficace una politica di conservazione dell'hardware, che ha senso solo in prospettiva museale, si può solo ricorrere a una costosa e impegnativa politica di costante e progressivo aggiornamento evolutivo dell'hardware, del software e dei formati dei dati. Sapendo che ad un certo punto i principi tecnologici su cui si basa la funzionalità della biblioteca si saranno comunque evoluti al punto che essa per continuare a esistere dovrà essere ricostruita da zero - come è ben esemplificato dal 'passaggio' dal motore a vapore al motore a scoppio; o dal motore a scoppio al motore elettrico.

2 A che cosa serve una biblioteca digitale?

Tendiamo a pensare che iniziative di ricerca costose e delicate debbano avere uno scopo forte, che ne giustifichi l'esistenza. La cosa non è ovvia: nelle scienze dure ci sono ambiti di ricerca (la fisica delle particelle, per esempio) che comportano spese immense a tempo indeterminato, in assenza di chiari scopi che configurino una prevedibile utilità di qualche genere per la società che li finanzia. Ciò però non suscita reazioni negative perché in quei campi la ricerca di base con tutti i suoi costi è ammessa senza limitazioni. Le scienze umane si trovano invece a dover giustificare nella società non solo lo scopo e l'utilità delle loro ricerche ma anche la soste-

nibilità economica degli investimenti che richiedono: di qui la domanda sugli scopi di una biblioteca digitale. Di primo acchito si può pensarli come lettura e ricerca di passi. Ma intraprendere la creazione di una biblioteca digitale semplicemente per permettere la lettura delle opere sullo schermo dei computer non sarebbe sostenibile. La disponibilità di opere in formato digitale (digitalizzate) permette però anche la ricerca e individuazione di specifici passi del testo: una semplice ricerca di parole permette di individuare il passo o i passi rilevanti per il proprio scopo, in modo simile a ciò che in passato si faceva consultando le concordanze di un'opera. Le concordanze erano uno strumento di difficile produzione e quindi realizzato solitamente per opere reputate di grande importanza; la digitalizzazione permette lo stesso tipo di lettura analitica per qualsiasi opera disponibile in formato digitale. Ma sono possibili anche altre ricerche più raffinate, come quelle che mirano a individuare la presenza di espressioni di uno scrittore nelle opere di altri scrittori successivi, ricerche in cui l'acume dello studioso viene supportato dalla velocissima, precisa ed esauriente ricerca effettuata dal computer; o quelle che cercano compresenze (co-occorrenze) di parole.

3 Attività consuete, attività nuove

Fin qui vengono realizzate nel mondo digitale le medesime attività che si possono realizzare anche nel mondo fisico con i testi a stampa - ma con maggiore precisione e soprattutto con maggiore velocità, il che significa che ricerche concepibili ma non realizzabili praticamente con le opere a stampa (un caso su tutti: ricerche di co-occorrenze) diventano invece possibili. Ma si possono concepire anche attività nuove, attività che sfruttano a fondo la natura digitale dei testi e li portano al centro dello sviluppo del web del futuro, il cosiddetto 'web semantico'. Perché nel mondo digitale che sta crescendo la cultura classica non deve per forza rimanere nella posizione specialistica/marginale che occupa attualmente nel mondo fisico, continuamente messa alla prova per dimostrare che serve a qualcosa, che ha qualcosa di dire che vale la pena di ascoltare, in quanto appare irrimediabilmente legata al passato - ma può e deve avere uno spazio in cui tutti la possano incontrare come parte del patrimonio culturale complessivo a disposizione nel mondo digitale. Perché sarebbe veramente paradossale che non ci fossero (anche) gli studiosi di scienze umane a far nascere e crescere il web semantico.

4 Licklider e Engelbart: *Augmenting Human Intellect*

Nel 1960 Joseph Carl Robnett Licklider scriveva:

Nella primavera-estate del 1957 [...] cercai di tener traccia di ciò che una persona moderatamente tecnica faceva effettivamente durante le ore che considerava dedicate al lavoro. [...]

Circa l'85 per cento del mio tempo di 'pensiero' era speso nel trovare una posizione per pensare, nel prendere una decisione, nell'imparare qualcosa che sapevo essermi necessario. Molto più tempo veniva speso nel trovare o ottenere informazioni che nello studiarle. Molte ore erano spese nel tracciamento di grafici e altre ore nell'insegnare ad un assistente come tracciarli. Quando i grafici erano finite le relazioni erano ovvie a colpo d'occhio, ma il tracciamento era stato necessario per renderle così evidenti.

Nel corso del periodo che esaminai, insomma, il mio tempo di 'pensiero' era stato dedicato principalmente ad attività che erano essenzialmente esecutive o meccaniche: cercare, calcolare, tracciare, trasformare, determinare le conseguenze logiche o dinamiche di un insieme di assunzioni o ipotesi, preparare la strada per una decisione o una visione approfondita. Per di più le mie scelte di che cosa fare o non fare erano determinate in misura imbarazzantemente grande da considerazioni di fattibilità pratica, e non di capacità intellettuale.

Il principale suggerimento fornito dalle scoperte appena descritte è che le operazioni che riempiono la maggior parte del tempo che si presume dedicato al pensiero tecnico sono operazioni che possono essere eseguite dalle macchine in modo più efficace che dagli uomini. Gravi problemi sono posti dal fatto che queste operazioni devono essere eseguite su differenti variabili e in sequenze impreviste e in costante mutamento. (Licklider 1960, 6)

Senza lasciarsi fuorviare dal fatto che il contesto in cui si colloca il lavoro di Licklider sia di tipo tecnico, il senso complessivo del suo discorso è chiaro e pertinente anche al lavoro di studio in ambito umanistico: il tempo dedicato ad attività eminentemente creative sul piano intellettuale è solo una piccola parte del tempo complessivo speso in una ricerca. La maggior parte è spesa in attività esecutive/ripetitive che però «devono essere eseguite su differenti variabili e in sequenze impreviste e in costante mutamento».

Nel 1962 Douglas Engelbart nell'introduzione di *Augmenting Human Intellect. A Conceptual Framework* scriveva:

Con 'potenziare l'intelletto umano' intendiamo accrescere la capacità di un uomo di affrontare una situazione che presenta un problema complesso, di acquisire comprensione adatta alle sue specifiche necessità, e

derivare soluzioni per i problemi. Capacità accresciuta, in questo senso, è usato per indicare una miscela dei seguenti elementi: comprensione più rapida, comprensione migliore, possibilità di acquisire un livello di comprensione utile in una situazione che prima era troppo complessa, soluzioni più rapide, soluzioni migliori, e la possibilità di trovare soluzioni per problemi che prima sembravano irrisolvibili. E con 'situazioni complesse' indichiamo i problemi professionali dei diplomati, degli uomini d'affari, di chi fa ricerca in sociologia, biologia, fisica, degli legali, dei progettisti - sia che la situazione problematica esista per venti minuti o per venti anni. Non parliamo di trucchi intelligenti e isolati che aiutano in una particolare situazione. Ci riferiamo invece ad un modo di vivere in un dominio integrato dove sensazioni, procedimenti per prove ed errori, intangibles, e la qualità specificamente umana della 'percezione della situazione' coesistono in modo utile con concetti potenti, terminologia e notazione efficace, metodi sofisticati e aiuti elettronici di alto livello. (Engelbart, 1962, 1)

In una sorta di continuazione ideale delle riflessioni di Licklider, Engelbart ritiene che i computer servano al lavoro intellettuale se si riesce a farli essere parte di «dominio integrato dove sensazioni, procedimenti per prove ed errori, e la qualità specificamente umana della 'percezione della situazione' coesistono in modo utile con concetti potenti, terminologia e notazione efficace, metodi sofisticati e aiuti elettronici di alto livello».

Quando un computer dotato di un programma apposito permette, per esempio, di produrre la concordanza di qualsivoglia testo in formato digitale siamo dentro il 'paradigma Licklider': attività esecutive e ripetitive che si basano su un input specifico vengono eseguite da un assistente che le espleta in modo efficientissimo lasciando allo studioso più tempo per ragionare sul significato degli esiti delle ricerche. Il 'paradigma Engelbart' suggerisce qualcosa di più, cioè che il computer diventi parte integrante dell'ambiente di lavoro intellettuale in cui le qualità propriamente umane (sensazioni, procedimenti per prove ed errori, *intangibles*, percezione della situazione) si caratterizzano per essere non-totalmente-formali, il che è un pregio perché permette di operare in modo efficace in tutte quelle situazioni in cui le informazioni, i dati, le conoscenze, non sono completi. In questa situazione il computer e gli esseri umani possono interagire grazie a operazioni esterne e automatiche sui simboli: ciò accade quando i simboli con cui l'essere umano rappresenta i concetti su cui opera possono essere ordinati sotto i suoi occhi, spostati, archiviati, richiamati, modificati, secondo regole anche estremamente complesse, in risposta rapidissima a un minimo di informazione fornita dall'essere umano stesso, per mezzo di speciali dispositivi tecnologici cooperativi:

Nei limiti di ciò che noi possiamo oggi immaginare, potrebbe trattarsi di un computer, con cui potremmo comunicare in modo rapido e facile, accoppiato con un display a colori tridimensionale all'interno del quale potrebbe costruire immagini estremamente sofisticate - con il computer capace di eseguire un'ampia gamma di elaborazioni su parti di quelle immagini, o sulle immagini intere, in risposta automatica a istruzioni umane. (Engelbart 1962, 25)

Engelbart nel 1962 pensava ai simboli come sorte di ideogrammi ma la funzione di quelli che egli chiama simboli la possono anche svolgere le parole, che non rappresentano più se stesse, ma che fungono da etichette, indicatori, che rappresentano gruppi di significati o di concetti - in modo tale per cui gli esseri umani e i computer cooperano nella costruzione di nuova conoscenza.

5 Dalla ricerca di parole alla ricerca di categorie: l'annotazione semantica

Chiunque abbia usato un computer dotato di un programma per ricerca di parole sa bene che se vuole studiare il concetto di abitazione e le forme in cui esso si manifesta nei testi la sua sarà una ricerca di molteplici parole, quale che sia la lingua: per indicare la casa di campagna, la casa piccola, la casa imponente, la casa pubblica, la casa privata, e altro ancora, altrettante e più saranno le parole effettivamente usate nel testo in esame. Meglio sarebbe se tutte le parole afferenti all'ambito lessicale desiderato potessero essere ricercate contemporaneamente passando da una 'analisi del testo quantitativa' a una 'qualitativa', dove 'quantitativo' è ciò che si limita alle stringhe, alle esatte sequenze alfanumeriche da trovare nei testi, mentre 'qualitativo' è ciò che mette in gioco la semantica.

Nel momento in cui in un contesto digitale si introduce la semantica, 'il gioco si fa duro': la semantica, il riconoscimento dei significati (ma si dovrebbe più propriamente dire la corretta classificazione dei significati), la disambiguazione, si ottengono pragmaticamente con simulazioni della conoscenza che un essere umano utilizzerebbe per svolgere quei compiti. Ma pur con le simulazioni, pur con un certo tasso di errore, si possono ottenere risultati interessanti.

Qui si completa il ragionamento partito dalla domanda 'adesso che abbiamo le biblioteche digitali, che cosa ne facciamo?', per la quale le attuali risposte non possono continuare a rimanere valide o per la troppo grande sproporzione fra l'impegno anche economico necessario per realizzarle e il risultato che si ottiene ('leggiamo i testi') o perché non sfruttano appieno le opportunità offerte dal digitale ('facciamo ricerche di parole'). Una delle

possibili risposte adeguate è 'annotiamo i testi per poter effettuare ricerche semantiche', cioè 'portiamo i testi nel web semantico': formulazioni che possono lasciare indifferenti, come se si trattasse di non-risposte alla domanda da cui si è partiti o di discorsi semplicemente tecnici che non toccano lo studioso. In realtà 'portare i testi nel web semantico' significa impiegare ontologie informatiche (esistenti, o da creare ad hoc) per (de) scrivere nel testo in modo formalizzato il senso di parole o proposizioni che esso contiene. Una volta effettuata questa annotazione, essa può essere utilizzata per fare ricerche nel testo per mezzo del computer. Un'ontologia infatti è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di conoscenza, organizzata in una gerarchia di categorie; tale strutturazione permette di costruire ed effettuare per mezzo del computer semplici ragionamenti di tipo formale/deduttivo. Ad esempio se si descrivono per mezzo di un'ontologia gli elementi della geografia fisica (monti, fiumi, mari, territori, ecc.), non solo si indicherà che il Po è un fiume; ma il fiume sarà connesso con altre 'acque dolci' come torrenti; ma anche laghi; e pure stagni; ma più alla lontana anche con altre acque come le 'acque salate' dei mari e degli oceani. Ciò permetterà di ottenere tra gli esiti di una ricerca i passi in cui si menziona il Po sia che si cerchino i 'fiumi', sia che si cerchino le 'acque dolci', sia che si cerchino le 'acque' in generale. E dunque l'annotazione semantica dei testi per mezzo di ontologie è una risposta adeguata alla domanda da cui si è partiti: apre prospettive di studio dei testi molto ampie e innovative, non praticabili senza le biblioteche digitali.

6 Un caso concreto: Geolat e la biblioteca digitale digilibLT

La riflessione teorica sin qui concisamente delineata rimanda in concreto all'attività di ricerca in corso presso il Dipartimento di Studi Umanistici dell'Università degli Studi del Piemonte Orientale con i progetti digilibLT e Geolat. Il primo è stato cofinanziato dalla Regione Piemonte con un bando del 2008 per la ricerca nelle scienze umane e sociali, mentre il secondo è stato finanziato dalla Fondazione Compagnia di San Paolo, in entrambi i casi al termine di procedure di *blind peer evaluation*.

6.1 digilibLT

DigilibLT, Digital Library of Late-Latin Texts (<http://www.digiliblt.uniupo.it>), è un progetto di ricerca di cui è responsabile Raffaella Tabacco, condiretto con chi scrive, per la creazione di una biblioteca digitale del latino tardo, limitata agli autori pagani. La biblioteca raccoglie alla data di composizione di questo scritto 131 opere liberamente disponibili

per la lettura e il download sotto licenza Creative Commons BY-NS-SA, in 4 formati: TXT, PDF, XML/TEI e ePub. Mentre per i formati TXT PDF e XML/TEI le finalità e gli usi sono chiari, può essere interessante segnalare che la disponibilità delle opere anche in formato ePub, unitamente alla possibilità di scaricarle tutte in un unico blocco, permette di portare con sé in un ebook reader tutta la letteratura latina tarda, potendo leggere qualsiasi passo di qualsiasi opera – ciò che praticamente nessuno studioso ha mai potuto realizzare fino ad ora. Per una descrizione ed analisi più dettagliata della biblioteca si veda il contributo di Borgna e Musso in questo stesso volume.

6.2 Geolat

Geolat, Geography for Latin Literature (<http://www.geolat.it>), è un progetto di ricerca di cui è responsabile chi scrive, per l'annotazione semantica dei nomi geografici nella biblioteca digitale digilibLT e in altre biblioteche digitali di testi latini; sui testi così annotati saranno poi possibili ricerche di testo che riguardino specificamente la conoscenza geografica contenuta nei testi stessi.

Il punto di partenza sono le opere degli scrittori latini disponibili in una biblioteca digitale. Si è scelto di partire lavorando sui testi di digilibLT sia per ragioni pratiche sia allo scopo di valorizzare la biblioteca che si sta costruendo, ma – e questo è molto importante – si può operare su qualsiasi opera digitalizzata presente in qualsiasi biblioteca purché la sua licenza lo consenta (occorre una licenza che definisca l'open access alle opere) – l'annotazione infatti può anche essere totalmente esterna al testo annotato, si parla in tal caso di *standoff markup*. In fase iniziale in Geolat i testi vengono copiati dalla fonte (digilibLT) e duplicati internamente, per dar luogo a una annotazione mista, in parte interna ai testi (*inline markup*), in parte esterna. Per avere un'idea di come si può presentare la parte di annotazione inline si veda il metaesempio qui sotto:

```
pax ita conuenerat ut Etruscis Latinisque fluuius <geogName
ref='http://www.geolat.it/geoDat/aTiber' xml:id='tiber01'>
Albula</geogName>, quem nunc <geogName ref='http://www.
geolat.it/geoData/Tiber' xml:id='tiber02'>Tiberim</geogname>
uocant, finis esset
```

Per l'annotazione di passi di testo contenenti riferimenti a luoghi ed entità geografiche lo standard TEI offre un vocabolario piuttosto ricco (capitolo 13 delle TEI Guidelines, «Names, Dates, People, and Places»). Nell'esempio viene usato il termine `geogName`, che “identifica un nome o una frase che si riferisce ad un luogo fisico” per il quale sono utilizzati i seguenti

attributi: @xml:id che identifica univocamente l'elemento annotato e il cui scopo è di permettere rinvii da parte di programmi ai passi di testo che menzionano l'entità geografica; @ref che contiene l'URI che identifica l'entità geografica citata nel passo.

Il dettaglio della descrizione analitica del nome di luogo basata sull'ontologia si collocherà nello standoff markup, utilizzando come connettore tra passo di testo e markup l'attributo 'ref' citato sopra.

L'ontologia permette di descrivere in modo formalizzato i luoghi antichi e la percezione che di essi aveva la cultura del tempo. Essa quindi da un lato contiene la strutturazione concettuale delle entità fisiche (acque, rilievi, e così via) e antropiche (confini, territori, capoluoghi/capitali, ecc.) del mondo antico, ma anche del mondo contemporaneo. Quest'ultimo aspetto potrebbe apparire superfluo o addirittura fuorviante, ma in realtà tenerne conto permette di individuare le corrispondenze e le relazioni tra le entità della geografia antiche e quelle della geografia attuale. Si pensi ai nomi dei monti o dei fiumi, o delle città: occorre poter indicare che la città che si chiamava Augusta Praetoria attualmente si chiama Aosta e continua ad avere la medesima posizione nello spazio; ma dall'altro essa affronta anche questioni complesse e sfaccettate perché occorre rendere conto tra l'altro del fatto che

- le città avevano un fondatore;
- le colonie avevano una madrepatria;
- molti luoghi naturali erano sacri a una qualche divinità;
- le divinità avevano una gerarchia
- dei e dee personificati, principi e qualità deificati (il dio Termine o la dea Giustizia), e altre divinità minori.

E queste caratteristiche possono cambiare da fonte a fonte, o possono presentarsi in modo a loro volta sfaccettato (a seconda dei testi e a seconda del tipo di discorso che viene sviluppato il fondatore di Roma è Romolo oppure anche Enea). Inoltre ci sono aspetti problematici quali:

- la misura dello spazio che ci giunge dal mondo classico è variabile da fonte a fonte e non certa in quanto non si possiedono corrispondenze esatte tra le unità di misura antiche e quelle attuali;
- l'identificazione e il posizionamento esatto dei luoghi possono essere ignoti o dibattuti, come nel caso delle isole Cassiteridi;
- la concezione complessiva dello spazio geografico è differente.

Tali differenze possono dar luogo a problemi irrisolvibili quando si tenti di organizzare e sistematizzare queste conoscenze. Nei testi latini poi si incontrano peculiarità espressive quali, ad esempio, l'indicare un territorio con il nome della popolazione che lo abita - i.e. funge da indicazione geo-

grafica un'espressione che non contiene un nome geografico. E per finire l'ontologia - in quanto viene progettata e costruita anche con finalità di studio (e non solo di descrizione) dei testi annotati - prevede l'esplicitazione di informazioni che si potrebbero definire di supporto, ad esempio

- le determinazioni di tempo relative all'uso di un determinato nome (come nel caso di Felsina/Bononia), o il periodo di tempo in cui un determinato territorio esiste con una identità specifica (la Dacia esiste come provincia a partire dal 102 d.C.);
- l'evoluzione nel tempo di un determinato luogo: Augusta Praetoria prima di essere una città era un castrum;
- o l'indicazione della fonte di autorità su cui si basa una determinata caratteristica associata al nome di luogo: ad esempio l'identificazione delle isole Cassiteridi con le isole Scilly viene completata dalla citazione bibliografica della fonte di tale identificazione; e se vi sono molteplici ipotesi più d'una può essere annotata.

Tali informazioni permettono poi in fase di ricerca sui testi di operare impiegando filtri specifici relativi al periodo in cui una data entità esisteva, o si presentava con determinate caratteristiche, o relativi alla fonte di informazione. In sostanza, ancora una volta, anche nel caso dell'ontologia vale il principio per cui lo strumento sarà tanto più utile quanto più si sarà progettato fin dall'inizio l'uso che se ne vuole fare o che se ne potrà fare.

L'ontologia geografica è accompagnata da un'ontologia testuale (Open Annotation), che permette di descrivere le relazioni tra autore, testo, suddivisioni del testo, passi, varianti, necessaria perché ogni entità geografica, man mano che le sue occorrenze vengono annotate nei testi, si arricchisce dell'elenco dei passi stessi. In tal modo l'ontologia non rimane una strutturazione della conoscenza, ma diventa una raccolta ordinata, enciclopedica, di informazioni.

Annotare i testi utilizzando l'ontologia è un'operazione complessa, che può essere fatta solo da studiosi perché occorre

- individuare i nomi di luogo (cioè riconoscerli come tali);
- disambiguare le omonimie (per esempio distinguere una dall'altra le varie "Alessandria");
- associare al nome del luogo le caratteristiche appropriate.

Tutte operazioni che richiedono solide conoscenze di lingua e cultura classiche. Nello stesso tempo è chiaro che servono anche strumenti per sveltire le operazioni di annotazione e a questo scopo si ricorre a un ambiente di lavoro specificamente destinato all'annotazione dei testi e a strumenti di lemmatizzazione integrati da procedure di Named Entities Recognition.

L'ambiente di lavoro permette di sveltire le operazioni ripetitive: se un

nome di luogo presente in un dato passo è già stato annotato in precedenza in un altro passo di un'altra opera, l'ambiente di lavoro lo segnala, lo studioso verifica che non si tratti di un'omonimia, e associa al passo di testo in esame la descrizione del nome di luogo già precedentemente inserita. La lemmatizzazione (che si realizza grazie alla collaborazione dell'Istituto di Linguistica Computazionale del CNR) e la NER permettono di individuare i nomi geografici, di identificarli, e di associare a essi un'annotazione minima che contiene alcune caratteristiche descrittive di base: le coordinate GPS e le traduzioni del nome nelle lingue moderne, ove disponibili. Lemmatizzazione e NER sono soggette a errori, tanto più sui nomi propri, ma forniscono un grande aiuto nell'annotazione dei nomi più frequenti, e permettono che il lavoro dello studioso si concentri sugli aspetti più creativi.

Ovviamente non è pensabile che un piccolo gruppo di persone possa provvedere a una così complessa annotazione dei nomi geografici in tutta la letteratura latina tardoantica in pochi mesi. L'intero sistema e ambiente di lavoro è pertanto concepito in modo da permettere la collaborazione di molteplici studiosi nel corso del tempo – perché in certa misura l'annotazione dei testi può essere concepita come un modo per lo studioso per riportare in contesto l'esito di uno studio. Se Mario Rossi pubblica un articolo in cui propone una nuova identificazione delle isole Cassiteridi, egli può annotare l'entità geografica 'isole Cassiteridi' in Geolat inserendo la sua identificazione nella geografia contemporanea, completa di citazione della fonte. Il testo antico diventa così contenitore della conoscenza che si possiede su di esso, in modo non molto dissimile da quanto avveniva con gli scholia.

Tutto questo complesso lavoro permette poi di costruire un'interfaccia di ricerca e consultazione dei testi geograficamente potenziata. Nella sua forma più estrema questo significa che sarà possibile fare ricerche sui testi partendo da una carta geografica sulla quale si traccia un cerchio, ottenendo in risposta l'elenco e la visualizzazione dei nomi di luogo di tale area presenti nelle opere e negli autori annotati, oltre all'elenco delle opere e degli autori che menzionano luoghi collocati in quell'area (ovviamente dagli elenchi dei passi si potrà andare alla loro lettura). Ma l'ontologia permette di creare ricerche o di filtrare quelle già fatte: 'la categoria delle acque negli scrittori del III e IV secolo d.C.'; oppure di creare ricerche miste che integrano la normale ricerca di parole con la ricerca di categorie dell'ontologia: 'la categoria dei rilievi e la menzione dei serpenti all'interno del medesimo capitolo nelle opere di genere storico'. In altre parole Geolat offre un'interfaccia di ricerca che aggiunge capacità geografiche a quelle consuete (ricerca di parole, generazione di concordanze, e simili); e le capacità geografiche hanno un certo grado di finezza semantica grazie all'impiego dell'ontologia.

L'interfaccia di ricerca qui rapidamente delineata è solo uno dei possibili

impieghi dell'annotazione del testo per mezzo dell'ontologia geografica. Infatti l'annotazione del testo sarà pubblicata sotto forma di dati aperti che qualsiasi servizio digitale potrà interrogare e riusare; per esempio un soggetto che opera nel mondo del turismo potrà mettere a punto la sua offerta cercando quali località antiche menzionate nei testi si trovano in una determinata area e individuandone i corrispondenti nomi attuali nell'ontologia di Geolat e nei testi latini; analogamente potrebbe fare un individuo che voglia costruire un viaggio sulle tracce della 'Peregrinatio Egeriae'; e analogamente potrebbe fare una persona semplicemente curiosa così come già accade con le mappe di Google: si clicca, si fa zooma, ci si sposta – come dire che il mondo antico è presente, è a portata di clic, ci si può curiosare dentro con lo stesso atteggiamento con cui si è curiosa nel mondo di oggi con Google Maps.

7 Per concludere

Quello che è stato qui velocemente delineato è un discorso molto complesso, in cui si connettono – al di là di una loro esplicita e consapevole intenzionalità – la ricerca e i suoi sviluppi, le finalità degli studiosi, l'interesse pubblico, gli sviluppi del web, lo spazio che negli anni a venire gli studi classici potranno avere o sapranno conquistarsi nel discorso pubblico che si volgerà nel mondo digitale. Fino a qualche anno fa gli studiosi potevano ignorare tutto ciò, concentrandosi sul fare buona ricerca, ma oggi non è più possibile. Una parte crescente delle fatiche degli studiosi va nella ricerca di finanziamenti, quei finanziamenti per i quali, sempre in misura più o meno grande, viene chiesto quale sia l'interesse pubblico a che quella data ricerca si realizzi – segno che le nostre discipline devono diventare capaci di raccontare in modo convincente nella società i loro scopi e i loro sogni, perché 'il classico' non può più far conto su un prestigio e un'autorevolezza riconosciute a priori.

Bibliografia

- Ciotti, Fabio; Lana, Maurizio; Tomasi, Francesca (2015). «TEI, Ontologies, Linked Open Data. Geolat and Beyond» [online]. *Journal Of The Text Encoding Initiative*. DOI 10.4000/jtei.1365.
- Corcione, Claudia et al. (2015). «GO! Geographical Ontology» [online]. URL <https://goo.gl/2WmQH8> (2017-11-13).
- Engelbart, Douglas (1962). *Augmenting Human Intellect. A Conceptual Framework* [online]. SRI Summary Report AFOSR-3223. URL <https://goo.gl/n5Y11E> (2017-11-13).
- Lana, Maurizio (2015). «Biblioteche digitali e Linked Open Data. Contenuti, connessioni, link, argomentazioni». *Digitalia*, 10, 20-9.
- Licklider, Joseph (1960). «IRE Transactions on Human Factors in Electronics» [online]. URL <https://goo.gl/vndAUs> (2017-11-13).