

ALIM ieri e oggi

Edoardo Ferrarini

Dipartimento di Culture e Civiltà, Università di Verona, Verona, Italia
edoardo.ferrarini@univr.it

Abstract. ALIM (Archive of the Italian Latinity of the Middle Ages) aims to provide free online access to all the Latin texts produced in Italy during the Middle Ages. For several centuries, in fact, Latin represented the only language in which, in addition to historical documentation, many of the major creations of thought, science, and literature of the Middle Ages were expressed. Even when national languages imposed themselves in written form, Latin never lost its role and prestige as transnational language. Born in the Nineties, the original purpose for ALIM project is to integrate the linguistic information which, for the previous centuries, is made available in the Dictionary of Medieval Italian Latinity, promoted by the U.A.N. (National Academic Union) under the guidance of F. Araldi and P. Smiraglia (*Latinitatis Italicae Medii Aevi lexicon*, Florence 2001). In addition to providing critically reliable texts for the use of philologists, historians of literature, historians of institutions, culture, and science of the Middle Ages, the ALIM project aims to produce useful material for a dictionary of European Medieval Latin. The ultimate goal of the latter project is to collect and unify the lexicons produced by all the European states for their own medieval Latin texts.

Abstract. Il progetto ALIM (Archivio della Latinità Italiana del Medioevo) intende offrire alla libera consultazione online tutti i testi in latino composti in Italia nel corso del medioevo. Il latino rappresentò, infatti, per molti secoli la sola lingua in cui ebbero espressione scritta molte delle principali creazioni del pensiero, della scienza e della letteratura del medioevo, oltre che la documentazione storica; anche quando si affermarono nella scrittura le lingue nazionali, il latino non perse mai la funzione e il prestigio di lingua transnazionale. Il progetto di ricerca, nato negli anni Novanta, si propone di completare le informazioni linguistiche reperibili, per i secoli precedenti, nel dizionario della latinità medievale italiana, promosso dall'U.A.N. (Unione Accademica Nazionale) e curato da F. Araldi e P. Smiraglia (*Latinitatis Italicae Medii Aevi lexicon*, Firenze 2001). Il progetto ALIM intende contribuire, oltre che a fornire testi criticamente affidabili per la consultazione di filologi, storici della letteratura, storici delle istituzioni, della cultura e della scienza medievali, anche a produrre materiali analizzabili per il progetto di un dizionario del latino medievale europeo, che integri i lessici delle rispettive latinità medievali elaborati da tutti gli Stati europei.

Keywords: Italia medievale, latino medievale, lessicografia mediolatina, corpora testuali, codifica XML-TEI

1 Introduzione

Il 1939 non merita certo di essere ricordato come un anno fausto per il progresso della civiltà umana. Eppure, per una di quelle curiose coincidenze che talvolta ci regala la storia, proprio nei momenti in cui si preparava il *Blitzkrieg* della Germania hitleriana contro la Polonia, veniva inaugurata, presso l'Università Cattolica di Milano, la prima cattedra universitaria italiana di Letteratura latina medievale. I motivi di tale singolare ritardo rispetto agli altri Paesi europei, pur meritevoli d'essere indagati, qui non ci interessano.¹

Inaugurava la disciplina un giovane studioso appena trentenne, Ezio Franceschini, che nella prolusione al suo corso, letta il 18 aprile del '39 di fronte alle autorità accademiche, volle delineare *Limiti e compiti di una nuova disciplina*.² Mettendo in guardia contro le storie letterarie e le grandi operazioni di sintesi, pericolose quando prive del fondamento dato da solidi lavori monografici e – cosa ancor più importante – da edizioni attendibili delle fonti,³ assegnava come compito principale agli studiosi della latinità medievale innanzitutto la pubblicazione critica dell'«estesissimo materiale inedito sepolto nelle nostre biblioteche grandi e piccole». E si domandava: «Sarà un duro lavoro da facchini? Ebbene, faremo i facchini per gli storici e gli esteti di domani» ([13]: 77). La direzione, allora indicata con tanta lucidità e sicurezza dal Franceschini, è ancora oggi quella seguita dalla piccola comunità italiana dei mediolatini, che riversano costantemente le loro energie (e i pochi fondi di ricerca di cui riescono ancora a disporre) nel duro “facchinaggio” della pratica filologica.

Mi piace descrivere il progetto ALIM (Archivio della Latinità Italiana del Medioevo) come nient'altro di più che la fedeltà intelligente a quella stessa vocazione e missione nell'era delle reti telematiche. ALIM è nato, infatti, con l'intento principale di offrire alla libera consultazione on-line tutti i testi latini composti in Italia e/o da autori italiani nel corso del medioevo, durante un arco cronologico che spazia attualmente dall'VIII al XV secolo.

Il progetto ha mosso i primi passi a metà degli anni Novanta, epoca di grande fermento nel campo delle *Digital Humanities*, epoca di sperimentazioni riflesse – per così dire – già nel titolo con cui si presentava allora il progetto, ossia: «Repertorio linguistico automatico dei testi italiani in lingua latina». Ha potuto contare talvolta, dal 1996 ad oggi, sulla qualifica di «Progetto di rilevante interesse nazionale» (PRIN). Vi hanno partecipato Unità di ricerca – eufemismo burocratico per “facchini” – degli atenei di Venezia “Ca' Foscari”, Verona, Milano Statale, Siena, Roma Tre, Napoli “Federico II” e “Suor Orsola Benincasa”, Palermo. Coordinatori nazionali sono stati, negli anni, i professori Giovanni Polara, Gian Carlo Alessio, Antonio De Prisco e Francesco Stella.

In un ambito, quello digitale, così esposto ai rischi dell'obsolescenza, in cui, per dirla con le parole dell'*eretico high-tech* Clifford Stoll ([24]: 137), le applicazioni «hanno vita breve come la lattuga del verduraio», quale può essere il senso di una nuova riflessione scientifica su un progetto che si avvia a compiere ormai il suo ventesimo anno d'età? Io credo che ALIM possa rivelarsi un interessante punto di osservazione proprio in ragione di questa sua

¹ Qualche risposta la si potrà ricavare dalla lettura di [16].

² Già pubblicata nell'*Annuario dell'Università Cattolica* di quell'anno accademico [12], la prolusione è stata ristampata in anni recenti, in copia anastatica con note autografe dell'autore [13].

³ «Lo studio del Medioevo deve essere ancora uno studio prevalentemente monografico: di singoli autori, di singole opere, di singoli generi, di singoli movimenti di pensiero. Chi vuole su pochi dati anticipare delle sintesi, faccia pure: ma non si lamenti poi se la sua costruzione correrà il rischio di avere la consistenza della nebbia o i fondamenti sull'arena» ([13]: 77-8).

lunga ed eccezionale durata, che coinvolge anche, in buona misura, i gruppi di ricerca delle varie sedi universitarie coinvolte. In altre parole, forse comincia a darsi oggi lo spazio e la possibilità – mi si passi l'espressione – per una storiografia dell'Informatica umanistica, o qualcosa che molto gli assomiglia.⁴

Questa prospettiva mi esime – credo – da una presentazione tradizionale e dettagliata del progetto e del sito web che ne raccoglie attualmente i risultati,⁵ dato che sono entrambi ben noti alla comunità scientifica e sono stati già presentati in numerose occasioni congressuali.⁶ Mi limiterò a dare un elenco di quelli che vari recensori hanno giudicato essere i punti di forza del progetto;⁷ passerò poi a mettere in luce gli aspetti problematici che la sua realizzazione ha comportato, senza tacere i limiti delle soluzioni trovate in corso d'opera, nonché i veri e propri errori commessi nella fase di prima progettazione; infine, presenterò lo sforzo compiuto negli ultimi due anni di attività da parte del gruppo di ricerca per giungere ad una versione 2.0 di ALIM, ormai matura e pronta per sostituire, entro la fine del 2016,⁸ la precedente release.

2 I punti di forza

Questo l'elenco sommario delle caratteristiche che maggiormente distinguono questa biblioteca digitale mediolatina da altri archivi simili presenti in rete:

1. Il fatto di costituire, *in primis*, una mappatura linguistica regionale del mediolatino. Il progetto, infatti, nato in ambito lessicografico, si proponeva e si propone, come fine ultimo, il completamento delle informazioni linguistiche reperibili nel dizionario della latinità medievale italiana, il *Latinitatis Italicae Medii Aevi lexicon* di Francesco Araldi e Pasquale Smiraglia (Bruxelles 1939-1964; *editio altera*: Firenze 2001). Della lingua latina in uso nel medioevo – va precisato – esistono lessici sia a carattere universale o specialistico, sia a carattere nazionale o regionale:⁹ sono questi ultimi, però, i più idonei a rappresentare la straordinaria mobilità ed articolazione del mediolatino, soprattutto a livello lessicale (si pensi, per fare un solo esempio, ai calchi linguistici dalle parlate volgari).
2. Il rispetto assoluto dell'edizione-fonte. I testi dell'archivio sono, infatti, riprodotti secondo l'edizione critica più recente, non coperta da diritti d'autore, ovvero secondo

⁴ Sul tema merita senz'altro la qualifica di pionieristico il saggio di Hockey 2004 [15]; mantiene un respiro internazionale [18], mentre può risultare utile, come agile sintesi, anche il cap. *Una breve storia dell'Informatica umanistica in Italia* in [26].

⁵ <<http://www.alim.df11.univr.it/>>.

⁶ La prima e più completa presentazione del progetto ALIM [1] avvenne nella cornice dell'VIII Incontro di Informatica umanistica, tenutosi presso l'Università degli Studi di Verona nel 1999 (cronaca del convegno e recensione del volume che ne raccoglie gli atti in [10]). Annunci e prospettive sull'avvio di una seconda fase del progetto non sono mancati nelle più recenti presentazioni, in sede scientifica, dell'archivio (mi riferisco, in particolare, a [22] e [23]).

⁷ Cfr., ad es., la recensione di [21], apparsa in *Reti Medievali*, rivista telematica di riferimento per la medievistica internazionale.

⁸ Il nuovo sito è già in rete all'indirizzo <<http://it.alim.unisi.it/>>: questa nuova versione dell'archivio ALIM è ora l'unica che viene aggiornata con l'inserimento di nuovi testi.

⁹ Cfr., per un primo orientamento, [25] e [9]: 265-7.

- quella più attendibile da un punto di vista filologico; i testi vengono mantenuti nella *facies*, anche tipografica, scelta dall'editore (numerazione delle pagine, suddivisione delle righe, caratteri speciali e convenzioni varie), in modo da poterli citare correttamente, secondo le consuetudini scientifiche proprie delle discipline linguistiche e letterarie.
3. La banca-dati ALIM è l'unica a comprendere edizioni critiche prodotte ex novo dallo stesso gruppo di ricerca. Mi limito a citare, per l'ultimo biennio, il *Liber inferni* di Domenico Bandini, a cura di Francesco Stella e Marcella Gambineri (*born-digital edition* 2014); l'opera del dettatore Maestro Guido, operante all'epoca del Barbarossa, per la cura di Elisabetta Bartoli (Firenze 2014); il ricchissimo *Epistolario* di Pier della Vigna, di cui l'équipe diretta da Edoardo D'Angelo ha pubblicato a stampa anche una prima traduzione integrale in lingua italiana (Soveria Mannelli 2014).
 4. L'archivio ALIM è l'unico a comprendere tanto fonti letterarie quanto fonti documentarie. Le grandi raccolte diplomatiche di area italiana,¹⁰ digitalizzate con lo stesso scrupolo e la stessa metodologia dei testi letterari, ma anche con gli opportuni adattamenti, arricchiscono la banca-dati dei testi più ricchi in assoluto di varianti grafiche e neologismi.
 5. Il motore di ricerca include uno strumento – perfezionabile certo – per la gestione delle allografie (*michi* per *mibi* o *nuncius* per *nuntius*).
 6. Riguardo alle tematiche oggetto del presente incontro di studi, penso si possa parlare di ALIM come di un caso di successo, per il dialogo riuscito e per la collaborazione instaurata negli anni tra il gruppo di ricerca e alcune delle istituzioni culturali impegnate nell'ambito del *Cultural Heritage*. E non mi riferisco solo, com'è ovvio, alle biblioteche e agli archivi documentari, pubblici e privati, che ci hanno fornito i testi oggetto del nostro trattamento digitale, ma anche all'«Unione Accademica Nazionale di Roma», prestigioso ente fondato nel 1923 e ora vigilato – come usa dire – dal MiBACT, e all'«Union Académique Internationale», con sede a Bruxelles, federazione accademica finalizzata a promuovere la collaborazione scientifica internazionale, enti – questi due – che hanno concesso sin dal primo momento il loro patrocinio al progetto.

3 Gli aspetti problematici

Come detto poc'anzi, ALIM ha una storia ormai ventennale ed ogni storia, in quanto tale, è un impasto di felici intuizioni, ma anche di inevitabili errori da cui imparare. Nel nostro caso, in particolare, si è trattato fin da subito di sperimentare la difficile integrazione fra competenze lessicografiche e filologiche, da una parte, e metodi e strumenti computazionali, propri dell'informatica umanistica, dall'altra; non è un caso, dunque, che i maggiori errori o leggerezze della prima fase progettuale di ALIM si siano evidenziati soprattutto negli aspetti riguardanti la codifica e il trattamento digitale dei testi.

¹⁰ Fra le raccolte già in tutto o in parte digitalizzate si trovano il *Codex diplomaticus Cavensis*, i *Documenti per la storia della città di Arezzo nel Medio Evo*, i *Regii Neapolitani archivi monumenta edita ac illustrata* (testimonianze preziose – queste ultime – in quanto gli originali pergamenei andarono distrutti durante l'ultimo conflitto mondiale); un'altra grande raccolta documentaria, di cui è programmata a breve la digitalizzazione completa, è il *Codice diplomatico Verginiano*, edizione delle carte conservate nell'archivio dell'abazia di Montevergine (Avellino), fondata all'inizio del XII secolo.

1. Bisogna ammettere, innanzitutto, che nei primi anni di sviluppo del progetto fu sottostimato, in generale, il tema del diritto d'autore. I testi digitalizzati in ALIM, infatti, riproducono l'edizione critica a stampa più recente o quella considerata filologicamente più attendibile: le edizioni critiche, com'è noto, sono il risultato dell'attività di revisione scientifica della tradizione manoscritta, di quella indiretta e delle precedenti edizioni (ove presenti), svolta da un soggetto, l'editore, che cerca di ricostruire nella sua integrità e forma originaria il testo, liberandolo da errori e arbitrarie deformazioni introdotte nel corso del tempo. Al testo così ricostruito si accompagnano generalmente una nota introduttiva e un commento critico (apparato), che spiegano e motivano puntualmente le scelte compiute nelle fasi di *recensio* e *constitutio textus*. ALIM si prefigge di dare la possibilità all'utente di «trasferire sul suo scrittoio il testo di cui ha necessità in una forma coincidente con quella che appare nella edizione-fonte a stampa: eccettuano gli apparati di varianti e di note al testo, omessi necessariamente, come si sa, per non essere tenuti ad entrare nel malcerto e scivoloso terreno della corresponsione dei diritti d'autore» [1]. L'evoluzione normativa, la giurisprudenza e le direttive europee,¹¹ tuttavia, hanno chiarito che l'editore – il filologo, cioè – è titolare del diritto d'autore non solo sull'apparato e sul corredo scientifico, ma anche sul testo stesso della sua edizione, tutelata in quanto frutto del lavoro intellettuale e di cui l'autore acquista la titolarità per il fatto della creazione. Il testo critico, ovviamente, dovrà presentare il carattere della creatività e originalità, distinguendosi dalle altre edizioni critiche dell'opera. Secondo il dettato attuale della legge, la tutela è accordata per una durata di venti anni, a partire dalla prima lecita pubblicazione. Inoltre, all'autore del testo critico è riconosciuto il diritto all'indicazione del nome sull'opera pubblicata, fermi restando i rapporti contrattuali con il titolare del diritto di utilizzazione economica, cioè con colui che provvede alla pubblicazione dell'opera (la casa editrice).
2. L'architettura proprietaria del data-base e del sito web. Nella prima versione, ALIM era, in sostanza, una base-dati di modello relazionale, accessibile liberamente tramite web e basata sull'architettura «IBM Lotus Domino», che integra in unico software gestione del data-base e interfaccia web. La scelta fu dettata dall'opportunità di appoggiarsi ad uno strumento solido, collaudato e ottimamente supportato, ma ha rivelato nel tempo tutte le sue controindicazioni.
3. Anche la scelta di utilizzare, al momento della codifica, un markup procedurale, come HTML, anziché dichiarativo, ha mostrato ben presto i suoi limiti: il corsivo – per fare un solo esempio – è una strategia tipografica utilizzata dai diversi editori con significati diversi, come diversi sono i mezzi utilizzati a stampa per riprodurre la struttura astratta o logica del testo (le sue “divisioni”).
4. Oltre alla navigazione pagina per pagina all'interno dell'opera scelta, all'utente di ALIM era consentito il download dell'intero testo in formato Microsoft Word. Non si può nascondere il fatto che, all'interno della comunità scientifica di riferimento (quella degli studiosi della lingua e della letteratura mediolatina), questa è stata una delle caratteristiche più apprezzate del primo ALIM: «La possibilità di scaricare agevolmente edizioni

¹¹ Si veda, in particolare, l'art. 5 della direttiva europea 2006/116/EC – *Directive of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights*; utile per inquadrare il problema all'interno dell'ordinamento giuridico nazionale, anche se necessariamente da aggiornare con la normativa più recente, [2].

critiche in formato Word, corredate degli appositi rimandi alla versione cartacea, costituisce poi un ulteriore punto di forza del progetto ALIM rispetto ad altri siti che forniscono raccolte di fonti medievali» ([21]: 151). Negli stessi anni, la scelta di un download in formato Microsoft Word era difesa con queste parole: «Offrire una disponibilità ampia di testi mediolatini è anche uno dei fini che il progetto si propone: è stata pertanto scelta la soluzione, economica, di offrire testi privi di codifiche, quindi ‘puliti’, ed immediatamente fruibili anche per scopi diversi da quello dell’analisi linguistica o letteraria condotta con strumentazione informatica» ([1]: 81). Queste considerazioni di ieri potrebbero risultare oggi, a buon diritto, quasi irricevibili, ma il peccato – è noto – va confessato ad alta voce per poter sperare in una piena guarigione. A ben guardare, gli errori, in questo caso, non sono uno, ma tre. Il primo è stato quello di preferire, ancora una volta, una codifica di tipo proprietario. Il secondo l’aver ceduto al vero e proprio mito delle tecnologie informatiche, proprietarie o meno, come “soluzioni economiche”. A vaccinarci contro tale superstizione basterebbero le riflessioni dell’indiscusso pioniere dell’applicazione dell’informatica allo studio dei testi letterari, il padre Roberto Busa ([3]: 83-4): «Non c’è macchina che eguagli il computer nel far perdere somme enormi in tempo reale»; «Fin dagli inizi infatti ho riscontrato i danni procurati da una malintesa estensione del fatto che il computer risparmia lavoro (...), ma è un non-senso immaginare che con il computer l’uomo possa ridurre i suoi tempi di analisi classificatorie e di fantasia creatrice».¹² I testi privi di codifica, poi, lungi dall’essere più “puliti”, sono invece più poveri e, soprattutto, inservibili, proprio in quanto privi di codifica, ad eventuali, successive analisi e trattamenti computazionali: e questo è il terzo errore.

4 ALIM 2.0

Passo, dunque, a presentare brevemente le acquisizioni ed i progressi più recenti del progetto, che ci hanno consentito l’allestimento del nuovo ALIM 2.0,¹³ non senza avere prima precisato che la vera sfida, prima ancora che quella legata alla progettazione del nuovo ambiente digitale, è stata quella di riflettere criticamente, individuare e mettere in opera opportune strategie per la conservazione e la migrazione dei dati e dei meta-dati dell’archivio, in un’ottica di accorta *digital continuity*.

- La direttrice principale dell’innovazione in questione si può riassumere così: dall’open access, ispirazione già presente nel primo ALIM, all’*open source*. Questo ha comportato l’abbandono definitivo e irreversibile della piattaforma proprietaria di gestione del database dell’archivio.

Non sarà inutile fornire al lettore, nelle righe che seguono, qualche ulteriore dettaglio tecnico sull’architettura informatica della nuova piattaforma ALIM 2.0. Essa è basata sul framework MURUCA,¹⁴ uno strumento composto da vari moduli, completamente interoperabili, che costituiscono la spina dorsale del sistema di gestione di testi, immagini e contenuti

¹² Si tratta di alcuni dei sorprendenti, arguti e spesso provocatori pensieri ed aforismi di cui è ricco lo zibaldone autobiografico *Dal computer agli angeli* (si leggono ai numeri 0348, 0350 e 0351).

¹³ <<http://it.alim.unisi.it/>>.

¹⁴ <<http://www.muruca.org/>>.

editoriali del portale «Archivio della Latinità Italiana del Medioevo». Muruca permette un completo disaccoppiamento tra la logica di storage dei dati e la fruizione e visualizzazione dei contenuti. Ognuno dei moduli di cui si compone Muruca espone uno o più servizi necessari per il processo di pubblicazione e fruizione della digital library.

Il modulo di backend è un'implementazione generata sulla base della piattaforma software per lo sviluppo di applicazioni web denominata LAMP stack (Linux, Apache, MySQL, PHP) e, in dettaglio, utilizza il framework Symfony¹⁵ per l'autenticazione alla base di dati e la gestione delle entità presenti sulla piattaforma. Il modulo di backend è da considerare come il principale punto di accesso per la pubblicazione dei testi e delle immagini e consente la creazione, modifica e pubblicazione di un'opera e dei relativi metadati. Il modulo di backend interagisce con il database MySQL mediante il framework Doctrine¹⁶ che fornisce un livello di astrazione tale per cui in futuro, qualora si rendessero necessarie delle migliorie architetturali, si potrà agevolmente sostituire il repository MySQL con un altro database senza intaccarne la logica applicativa.

Per lo sviluppo del modulo CMS (*Content Management System*), che si occupa della pubblicazione e della fruizione dei contenuti editoriali, è stato utilizzato il molto diffuso WordPress.¹⁷ Mediante un semplice accesso con credenziali è possibile accedere al pannello di amministrazione ed intervenire sui vari componenti del portale di ALIM. Il modulo CMS utilizza come storage un database MySQL.

Per l'indicizzazione a faccette è stato personalizzato il motore di ricerca Solr, strumento standard e di buona affidabilità. Tutti i metadati inseriti tramite le schede del backend vengono processati e indicizzati opportunamente da Solr,¹⁸ allo scopo di rendere più veloci e flessibili le ricerche effettuate sui testi dell'archivio. Solr richiede come prerequisito la presenza di una JVM (*Java virtual machine*) che è stata opportunamente installata e configurata sul server. L'aggiornamento dell'indice di Solr avviene in tempo reale rispetto alla modifica dei metadati nel modulo di backend, ma, all'occorrenza, l'indice può essere aggiornato anche in modalità off-line, ricostruendone tutte le voci mediante opportuni script. Per la conversione degli originari documenti TEI-XML in altri formati, viene utilizzato OxGarage,¹⁹ un servizio web RESTful che permette di gestire la trasformazione di documenti in differenti formati. OxGarage è usato da Muruca proprio come servizio e il suo utilizzo permette la conversione dei documenti con codifica TEI-XML nei differenti formati di esportazione consentiti all'utente (TXT, PDF, HTML).

Il modulo di frontend, per finire, è composto da un'applicazione realizzata con il framework AngularJS.²⁰ L'applicazione interagisce con i moduli mediante uno strato di API; nell'architettura del sistema esistono due differenti API endpoint, uno esposto dal motore di ricerca Solr, un secondo, invece, dal modulo di backend. Il frontend, quindi, si limita a ricevere le informazioni memorizzate su tali storage ed a fornirne un'opportuna visualizzazione.²¹

¹⁵ <<https://symfony.com/>>.

¹⁶ <<http://www.doctrine-project.org/>>.

¹⁷ <<https://it.wordpress.com/>>.

¹⁸ <<http://lucene.apache.org/solr/>>.

¹⁹ <<http://www.tei-c.org/oxgarage/>>.

²⁰ <<https://angularjs.org/>>.

²¹ Per l'allestimento della nuova piattaforma ALIM 2.0 *open source*, nonché per la delicata fase di migrazione dei dati dell'archivio nel nuovo ambiente, il gruppo di ricerca ha potuto avvalersi della frut-

- Per quanto riguarda la codifica dei testi, ALIM 2.0 completa il passaggio allo standard XML-TEI P5,²² anche in vista di un'apertura alla collaborazione/integrazione con altri archivi testuali, banche-dati o dizionari. Questa scelta non ha riguardato solo i testi di nuova immissione, bensì anche tutti quelli già presenti in archivio (circa 350 opere della letteratura latina medievale e più di un migliaio di testi documentari), i quali sono stati migrati sulla nuova piattaforma, con grande sforzo, solo dopo averli adeguati al nuovo schema di codifica. Principio guida di tutta l'operazione è stata la ricerca di quella 'onestà del sorgente', su cui tanto insiste Tito Orlandi.²³

Un'apposita sezione del nuovo sito ALIM («Documentazione») dà accesso, oltre che ad un *Manuale* rivolto all'utente del sito, che indica come visualizzare i testi ed usare gli strumenti di ricerca e analisi testuale, all'intera documentazione relativa alla codifica. Vi trovano spazio: lo schema di codifica vero e proprio, in formato Relax NG; un documento di testo che presenta le specifiche convenzioni di codifica utilizzate (manuale di formazione per i collaboratori del progetto); un template del TEI Header ALIM, utilizzato come modello dai codificatori. Anche a proposito del markup, si deve sempre considerare il principio-guida, già presentato, di quest'iniziativa scientifica, ossia il rispetto assoluto, nella riproduzione digitale, dell'edizione-fonte. ALIM non rappresenta, infatti, in primo luogo, una raccolta di autori/opere, bensì una collezione digitale di edizioni critiche a stampa, in cui i testi, privati – come detto – degli apparati critici, vengono mantenuti nella *facies*, anche tipografica, adottata dall'editore.

Da questi presupposti derivano, nello schema di codifica, l'utilizzo del tag <pb> (page break), al fine di conservare la numerazione delle pagine, e <lb> (line break), a segnare la suddivisione delle righe all'interno della singola pagina a stampa, in modo da poter citare correttamente i testi in accordo con le consuetudini scientifiche proprie delle discipline linguistiche e letterarie. Per quanto riguarda le opere in versi, evidentemente, i marcatori utilizzati sono <l> (verse line) e <lg> (line group). La struttura gerarchica del testo è codificata mediante <div> (text division), senza numerazione, e <p> (paragraph). Particolare

tuosa collaborazione dell'azienda informatica Net7, con sede a Pisa (<<http://www.netseven.it/>>), azienda in possesso ormai di un lungo e prestigioso *curriculum* nel campo delle Digital Humanities e del Semantic Web. Un particolare ringraziamento, per la loro competenza e professionalità, mi sia consentito a Chiara Aiola, Francesca Di Donato e Romeo Zitarosa.

²² *TEI P5: Guidelines for Electronic Text Encoding and Interchange* [27]. Come introduzione generale al tema della codifica elettronica dei testi in ambito umanistico possono valere i lavori di [20] e [28] (in particolare al cap. 4: *La testualità digitale: i linguaggi di markup*, 103-142); di maggiore impegno è la prospettiva di [11] e [19].

²³ Lo studioso non esita a definire il sorgente come «il cuore di un'edizione digitale» ([17]: 86); e così si esprime: «Abbiamo già accennato a che cosa intendiamo per sorgente. Esso prima di tutto è ciò che in termine tecnico si chiama un file e come tale è gestibile solo dal computer e non ricade sotto i sensi umani. È utile immaginarlo come un'unica linea 'kilometrica' in cui sono registrati in sequenza i byte o gruppi di byte che corrispondono agli elementi discreti individuati dall'editore nel testo, distinto nei suoi diversi strati testuali a loro volta connessi con i diversi sistemi che lo hanno generato: semantico, linguistico, fonetico, grafico, ecc. Esso costituisce la rappresentazione digitale del testo, depositata nella memoria del computer, indipendentemente da come si manifesti quando è fatta apparire sullo schermo o sulla carta» (ivi: 88). Oltre alle ricerche di Orlandi, i ss. altri studi costituiscono l'orizzonte teorico delle scelte di codifica più oltre esplicitate: [14], [8], [4], [5] e [6], [7].

attenzione è posta alla disambiguazione del corsivo tipografico, che ha significati differenti a seconda delle scelte operate dai singoli editori: i marcatori utilizzati, in questo caso, sono principalmente <quote> (quotation), <title> (title) e <foreign> (foreign). Sono stati previsti differenti e progressivi livelli di codifica, tutti compatibili con lo schema di base: un livello 0, corrispondente all'esito della migrazione semi-automatica dalla versione precedente dell'archivio; un livello 0.5, che assicura la completezza dei metadati e la loro corretta memorizzazione; un livello 1, in cui anche la struttura del testo è rappresentata esaustivamente; un livello 2, infine, applicabile come sperimentazione ai testi ancora inediti. Nella sezione <revisionDesc> (revision description) del TEI Header si troverà l'informazione sull'ultimo livello di codifica applicato al singolo testo.

1. L'estensione delle funzioni di *text retrieval*. Attualmente la ricerca è consentita, attraverso un sistema di filtri a faccette, per autore, per titolo dell'opera, per tipologia testuale (prosa, versi, prosimetro), per genere letterario, per data di composizione. La ricerca *full-text*, invece, può avvalersi dell'uso di caratteri jolly, operatori booleani (and, or, not) e indici di prossimità.
2. La progettazione e l'implementazione di strumenti per l'analisi computazionale dei dati linguistici (come la produzione di indici e concordanze per singola opera o per una selezione di testi e, in prospettiva, la lemmatizzazione e la mappatura semantica dell'intero archivio) è resa possibile attraverso l'integrazione in ALIM 2.0 dell'applicazione *Lexicon*, sviluppata dal Centro Studi Comparati "I Deug-Su" dell'Università di Siena.²⁴ È facile profezia immaginare che sarà proprio questo, in futuro, il campo di ricerca e sperimentazione destinato al prossimo, auspicabile salto di qualità.
3. Un'assoluta novità è rappresentata, poi, dalla sezione del sito denominata «Collezioni». Si tratta, in breve, di moduli pensati soprattutto per la didattica universitaria: una scheda redazionale introduce un tema linguistico, letterario o storiografico, che viene poi sviluppato fornendo al lettore i collegamenti diretti ai testi dell'archivio che lo possono illustrare.²⁵

Ancora ad uno stadio aurorale, ma imprescindibile, è, infine, la riflessione più complessiva su quale potrebbe essere l'impatto pubblico o l'utilità sociale – per esprimersi così – del patrimonio culturale immateriale che si conserva e si rivela nell'ingente *corpus* testuale rappresentato da ALIM, quasi una sorta di monumentale "Edizione nazionale" digitale del patrimonio mediolatino. Oltre alle possibili utilizzazioni dell'archivio in ambito propriamente ecdotico (come l'identificazione di fonti altrimenti irreperibili nelle edizioni a stampa oppure il possibile impiego nella costruzione di apparati di fonti per nuove edizioni critiche), esso consente – crediamo – una fruizione anche per scopi diversi, raggiungendo un pubblico maggiore del bacino di origine del progetto. La mappatura linguistica e lessicale del

²⁴ <<http://www.lexicon.unisi.it/public/>>.

²⁵ Queste le prime collezioni allestite e già leggibili nelle pagine web di ALIM: *De historia* (sui tanti modi di narrare la storia che conosce il medioevo), *De legibus* (dedicata al diritto), *De civitate Aretina* (un percorso tematico sulla storia della città di Arezzo nel medioevo), *De itineribus* (viaggi, pellegrinaggi e missioni), *De natura rerum* (la sezione dedicata alle scienze naturali), *De medicina* (molte le fonti che ruotano attorno a Salerno, centro rinomato, nel medioevo, per questo tipo di studi) e *De dictamine sive de epistolis* (collezione di testi legati all'epistolografia e alla retorica medievale).

mediolatino, che ne era l'obiettivo iniziale, ha portato, infatti, all'allestimento di una vera e propria enciclopedia di contenuti, uno specchio della cultura, della spiritualità, delle idee, dei paesaggi, degli oggetti quotidiani persino, del medioevo occidentale, traggurati sempre attraverso la lente della produzione letteraria mediolatina e dei suoi generi.

Provvisoria conclusione storiografica e morale della favola: negli ultimi vent'anni Internet è finalmente uscita dal *suo* medioevo,²⁶ ma il *nostro* medioevo continuerà ad inseguirla.

References

1. Alessio, Gian Carlo. "Il progetto ALIM (Archivio della Latinità Italiana del Medioevo)." In *Biblioteche elettriche. Letture in Internet: una risorsa per la ricerca e per la didattica*, a cura di Francesco Santi, 73-81. Vol.1. Firenze: SISMEL - Edizioni del Galluzzo, 2003.
2. Basile, Massimo. "Le edizioni critiche e scientifiche." *AIDA. Annali italiani del diritto d'autore, della cultura e dello spettacolo*, 7 (1998): 14-26.
3. Busa, Roberto. *Dal computer agli angeli. Milleduecentosessantuno momenti di pensiero su come dopo 60 anni di avanguardia nell'informatica testuale io inquadrò le reti elettroniche entro quelle degli spazi della vita*. Vol. 2. Castel Bolognese: Itaca, 2001.
4. Buzzetti, Dino. "Digital Representation and the Text Model." *New Literary History* 33, num. 1 (2002): 61-87.
5. Buzzetti, Dino. "Diacritical Ambiguity and Markup." In *Augmenting Comprehension: Digital Tools and the History of Ideas. Proceedings of a conference at Bologna, 22-23 september 2000*, a cura di Dino Buzzetti, Giuliano Pancaldi e Harold Short, 175-188. London-Oxford: Office for Humanities Communication, 2004.
6. Buzzetti, Dino. "Oltre il rappresentare: le potenzialità del markup." In *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*, a cura di Lorenzo Perilli e Domenico Fiormonte, 39-62. Firenze: Le Lettere, 2011.
7. Ciotti, Fabio. "La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi." In *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*, a cura di Lorenzo Perilli e Domenico Fiormonte, 63-90. Firenze: Le Lettere, 2011.
8. Cover, Robin e Peter Robinson. "Encoding Textual Criticism." *Computers and the Humanities* 29, num. 2 (1995): 123-136. <http://dx.doi.org/10.007/BF01830706>.
9. De Prisco, Antonio. *Il latino tardoantico e altomedievale*. Vol. 23. Roma: Jouvence, 1991.
10. Ferrarini, Edoardo. "Biblioteche elettriche. Letture in Internet: una risorsa per la ricerca e per la didattica." *Studi medievali*, num. 50 (2009): 958-65.
11. Fiormonte, Domenico. *Scrittura e filologia nell'era digitale*. Torino: Bollati Boringhieri, 2003.
12. Franceschini, Ezio. "Limiti e compiti di una nuova disciplina." *Annuario della Università Cattolica del Sacro Cuore*, 1939.
13. Franceschini, Ezio. *Limiti e compiti di una nuova disciplina. Profilo letterario del Medioevo latino*, a cura di Claudio Leonardi e Francesco Santi. Vol. 9. Quaderni di cultura mediolatina. Spoleto: Centro italiano di studi sull'alto medioevo, 1993.
14. Gigliozzi, Giuseppe, a cura di. *Studi di codifica e trattamento automatico dei testi*. Roma: Bulzoni, 1987.
15. Hockey, Susan. "The History of Humanities Computing." In *A Companion to Digital Humanities*, a cura di Susan Schreibman, Ray Siemens e John Unsworth, 3-19. Oxford: Blackwell, 2004. <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1>.

²⁶ Quello di cui scriveva [29].

16. Leonardi, Claudio. "La filologia mediolatina 1944-1992." In *Enciclopedia Italiana di scienze, lettere ed arti*. Roma: Istituto della Enciclopedia Italiana, 1992.
17. Orlandi, Tito. "Teoria e prassi di una edizione computazionale." In *Digital Philology and Medieval Texts. Proceedings of the Arezzo Seminar 2006, 19-21 January*, a cura di Arianna Ciula e Francesco Stella, 85-92. Pisa: Pacini, 2007.
18. Orlandi, Tito. "Per una storia dell'informatica umanistica." In *Dall'Informatica umanistica alle culture digitali. Atti del Convegno di studi in memoria di Giuseppe Gigliozzi (Roma, 27-28 ottobre 2011)*, a cura di Fabio Ciotti e Gianfranco Crupi. Quaderni DigiLab, num. 1 (2012): 49-102. DOI: <http://dx.doi.org/10.7357/DigiLab-32>.
19. Pierazzo, Elena. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham, Surrey, England: Ashgate, 2015.
20. Renear, Allen Hume. "Text encoding." In *A Companion to Digital Humanities*, a cura di Susan Schreibman, Ray Siemens e John Unsworth, 218-239. Oxford: Blackwell, 2004. http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-3-5&toc.depth=1&toc.id=ss1-3-5&brand=9781405103213_brand.
21. Russo, Luigi. "ALIM, Archivio della latinità italiana del Medioevo." *Reti Medievali Rivista* 6, num. 1 (2005): 149-51. DOI: <http://dx.doi.org/10.6092/1593-2214/181>.
22. Stella, Francesco. "Encoding Issues in Philological Digital Editions." *Horizons: Seoul Journal of Humanities* 4, num. 2 (2013): 239-254.
23. Stella, Francesco. "Il problema della codifica nelle edizioni filologiche digitali." In *Nel segno del testo. Edizioni, materiali e studi per Oronzo Pecere*, a cura di Lucio Del Corso, Franco De Vivo e Antonio Stramaglia, 347-360. Firenze: Gonnelli, 2015.
24. Stoll, Clifford. *Confessioni di un eretico high-tech. Perché i computer nelle scuole non servono e altre considerazioni sulle nuove tecnologie*. tradotto da Andrea Antonini. New York: Doubleday, 1999.
25. Stotz, Peter. *Il latino nel Medioevo. Guida allo studio di un'identità linguistica europea*, a cura di Luigi Ricci. Vol. 1. Firenze: SISMEL - Edizioni del Galluzzo, 2013.
26. Tammaro, Anna Maria. "L'Informatica Umanistica e l'interdisciplinarietà." In *Umanisti e risorse digitali*, a cura di Anna Maria Tammaro. Parma: UniPR Co-Lab, 2012.
27. TEI Consortium. "TEI P5: Guidelines for Electronic Text Encoding and Interchange." *Text Encoding Initiative*, Dicembre 15, 2016. <http://www.tei-c.org/Guidelines/P5/>.
28. Tomasi, Francesca. *Metodologie informatiche e discipline umanistiche*. Vol.59. Roma: Carocci, 2008.
29. Zorzi, Andrea. "Il Medioevo di Internet. Lo stato delle risorse telematiche per gli studi medievali." *Quaderni medievali* 45 (1998): 146-179.

Ultima consultazione URLs: 01/02/2017