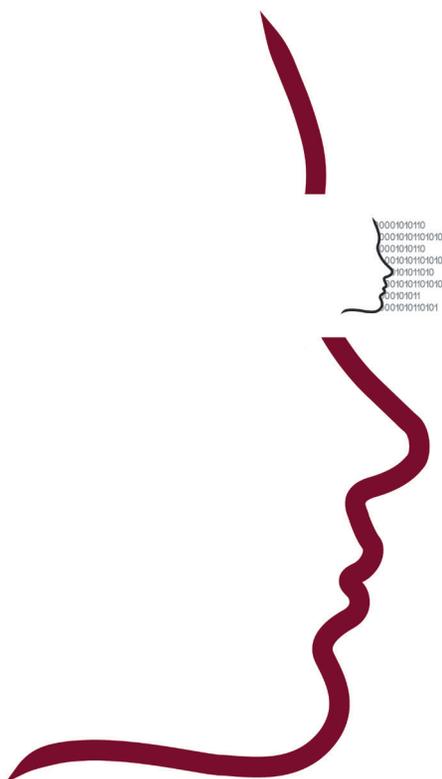


Digital Humanities

Progetti italiani ed esperienze di convergenza multidisciplinare

a cura di
Fabio Ciotti



Collana Convegni 24

STUDI UMANISTICI
Serie Quaderni DigiLab

Digital Humanities:
progetti italiani ed esperienze
di convergenza multidisciplinare

Atti del convegno annuale dell'Associazione
per l'Informatica Umanistica e la Cultura Digitale (AIUCD)
Firenze, 13-14 dicembre 2012

a cura di
Fabio Ciotti



SAPIENZA
UNIVERSITÀ EDITRICE

2014

Copyright © 2014

Sapienza Università Editrice

Piazzale Aldo Moro 5 – 00185 Roma

www.editricesapienza.it

editrice.sapienza@uniroma1.it

Iscrizione Registro Operatori Comunicazione n. 11420

ISBN 978-88-98533-27-5

DOI 10.13133/978-88-98533-27-5



Quest'opera è distribuita con licenza Creative Commons 3.0
diffusa in modalità *open access*.

Distribuita su piattaforma digitale da:

digilab

Centro interdipartimentale di ricerca e servizi
Settore Publishing Digitale

In copertina: Logo dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD).
Divieto di riproduzione.

Indice

Introduzione	1
PARTE I - DIGITAL HUMANITIES: VALUTAZIONE E INFRASTRUTTURE	9
1. La ricerca, la valutazione e la diffusione dei risultati nell'informatica umanistica. Problematiche aperte <i>Tito Orlandi</i>	11
2. La valutazione della ricerca umanistica: tra peer-review e bibliometria <i>Chiara Faggiolani, Giovanni Solimine</i>	15
3. Biblioteche digitali tra modellazione, gestione e valutazione <i>Maristella Agosti, Nicola Ferro, Gianmaria Silvello</i>	33
4. Livelli descrittivi, relazioni e contesti di produzione nella Sapienza Digital Library <i>Maria Guercio, Cecilia Carloni</i>	51
PARTE II - ARCHIVI DIGITALI: ARCHITETTURE E INTERFACCE	71
5. Il ruolo dell'Information Visualization nella progettazione di interfacce per archivi digitali eterogenei <i>Michele Mauri, Paolo Ciuccarelli</i>	73
6. Convergere a valle. Lo studio del punto di vista degli utenti degli ambienti culturali digitali e l'esperienza del progetto "Una Città per gli Archivi" <i>Pierluigi Feliciati</i>	89

7. Una risorsa online per la storia dell'arte: il database della Fondazione Federico Zeri <i>Francesca Mambelli</i>	113
PARTE III - FILOLOGIA DIGITALE; ESPERIENZE E SPERIMENTAZIONI	127
8. Stemma codicum: analisi e generazione semi-automatica <i>Gioele Barabucci, Angelo Di Iorio, Fabio Vitali</i>	129
9. Il progetto Visionary Cross: verso un'edizione digitale multimediale e distribuita <i>Roberto Rosselli Del Turco</i>	147
10. Many witnesses, many layers: the digital scholarly edition of the Iudicium coci et pistoris (Anth. Lat. 199 Riese) <i>Paolo Monella</i>	173
PARTE IV - BIBLIOTECHE DIGITALI TESTUALI; PROGETTI E SERVIZI INNOVATIVI	207
11. Immaginando il futuro digitale della traduzione letteraria: la traduzione dell'edizione 1860 di Leaves of Grass per il Walt Whitman Archive <i>Caterina Bernardini</i>	209
12. La localizzazione del Perseus Project in lingua italiana <i>Federico Boschetti</i>	221
13. Il progetto ENArC. Attività didattiche innovative e creazione di archivi digitali <i>Maria Rosaria Falcone</i>	235
14. Letteratura latina e ebook. Una prospettiva della biblioteca digitale digilibLT <i>Maurizio Lana</i>	247
15. L'ontologia BIA-Net: una base per la ricerca di informazioni secondo rapporti di rilevanza nella Bibliotheca Iuris Antiqui <i>Marco Giunti, Giuseppe Sergioli, Giuliano Vivonet</i>	263

Introduzione

Dino Buzzetti¹, Fabio Ciotti²

A un anno dalla sua costituzione, l'Associazione per l'Informatica Umanistica e la Cultura Digitale ha organizzato a Firenze nei giorni 13 e 14 dicembre 2012 il suo primo convegno, di cui qui si pubblicano gli atti. I lavori del convegno si sono articolati in tre sessioni, ciascuna delle quali ubbidiva a una precisa esigenza di intervento e di impegno prioritario per l'attività dell'Associazione. Delle tre sessioni proposte, le prime due sono state organizzate allo scopo di promuovere la discussione su due temi giudicati di particolare rilevanza nell'attuale fase di sviluppo delle cosiddette *Digital Humanities*. Come favorire la convergenza di competenze diverse per sostenere il processo, oramai avviato e sicuramente irreversibile, di transizione al digitale dell'intero patrimonio culturale? Come ripensare e riaffermare il ruolo essenziale che l'informatica applicata alle discipline umanistiche è responsabilmente chiamata a svolgere in questo epocale processo di trasformazione culturale? E a fianco di ciò, come riaffermare l'importanza culturale e scientifica della ricerca condotta nel campo dell'informatica umanistica e come accreditare e avvalorare i risultati prodotti in forma digitale con strumenti adeguati di valutazione e di certificazione scientifica?

Alla prima questione porta la constatazione della sempre crescente pervasività del digitale e del ruolo specifico che in questo contesto sono chiamati a svolgere gli studiosi dell'area umanistica. Da un lato, infatti, l'informatica interviene nel campo delle conoscenze tradizionali e delle stesse pratiche manuali, ricollocandole in una prospettiva affatto nuova che considera gli oggetti materiali primariamente come veicoli di informazione e che mira a cogliere in essi le tracce consolidate delle

¹ Università di Roma Tor Vergata. e-mail: fabio.ciotti@uniroma2.it.

² Presidente AIUCD e Fondazione per le Scienze Religiose Giovanni XXIII - Bologna. e-mail: dino.buzzetti@gmail.com.

relazioni con l'ambiente in cui sono stati prodotti e in cui vengono usati, oltre che le opportunità delle loro possibili interazioni in nuovi contesti e in nuove situazioni. Ma tale approccio non può ovviamente prescindere dalle competenze e dalle abilità specifiche tramandate dal sapere pratico e teorico tradizionale, col quale deve e intende necessariamente interagire [Rosner, Roccetti e Marfia]. Dall'altro lato, invece, le discipline umanistiche – ossia, nel loro insieme, quelle che August Boeckh concepiva come la «conoscenza di ciò che è ed è stato conosciuto (die Erkenntnis des Erkannten)» – nel rispondere al proprio «obbligo fondamentale» e nel dedicarsi al proprio impegno primario – la «preservazione della memoria culturale» [McGann] – non possono assolutamente ignorare le potenzialità che le nuove tecnologie informatiche offrono per l'archiviazione, la reperibilità e la valorizzazione dell'eredità culturale. E ancora meno possono sottovalutare le possibilità di un rinnovamento metodologico delle pratiche di ricerca generate dall'applicazione nel loro campo specifico delle procedure e dei modelli computazionali.

Nella progettazione e nella costruzione delle nuove raccolte digitali, si assiste così al convergere di diverse competenze, quelle biblioteconomiche sui sistemi di catalogazione e di organizzazione della conoscenza, quelle informatiche sulle tecnologie del Web, ora specificamente rivolte alla rappresentazione semantica del contenuto informativo delle risorse digitali e, infine, quelle disciplinari specifiche degli studiosi del patrimonio artistico letterario e culturale. Ed è proprio nel segno di questa convergenza che è stata pensata e organizzata la prima sessione del convegno, testimoniata qui dal saggio di Maristella Agosti, Nicola Ferro, Gianmaria Silvello, Biblioteche digitali tra modellazione, gestione e valutazione, dedicato a indicare i fondamenti per la progettazione, implementazione ottimale e valutazione dei sistemi informativi per gli oggetti culturali.

Si colloca a pieno titolo in questo contesto tematico il lavoro di Maria Guercio e Cecilia Carloni Livelli descrittivi, relazioni e contesti di produzione nella Sapienza Digital Library, sebbene l'intervento sia stato presentato nella sessione dedicata ai progetti.

Oggetto dell'articolo è infatti l'importante esperienza di progettazione e realizzazione di una piattaforma per la gestione e pubblicazione di risorse digitali, realizzata presso il Centro interdipartimentale di ricerca e servizi Digilab della Sapienza (digilab.uniroma1.it), uno dei maggiori

centri di Digital Humanities nel panorama italiano. L'intervento descrive l'architettura logica di questa infrastruttura ed evidenzia come questa preveda una modellizzazione dei contenuti che ne documenti sia il contesto di produzione sia le complesse relazioni strutturali e semantiche, mediante una descrizione a più livelli dei complessi documentari.

Il discorso sulle infrastrutture, peraltro, fornisce l'occasione per la proposta di una comune e convergente agenda di ricerca per le diverse comunità di studio e professionali coinvolte nella conservazione e valorizzazione del patrimonio culturale. La nostra Associazione è priva di una base istituzionale specifica nell'organizzazione disciplinare dell'insegnamento universitario, così come manca di una base professionale radicata nell'ambito di enti e organizzazioni specificamente dedicate all'impiego delle tecnologie informatiche per la creazione e il mantenimento di oggetti digitali prodotti con finalità eminentemente culturali. D'altra parte le competenze disciplinari teoriche e applicative proprie ed esclusive dell'informatica umanistica costituiscono una componente assolutamente necessaria e insostituibile in ogni azione diretta alla conservazione del patrimonio culturale o alla realizzazione di nuove forme di espressione artistica o culturale di natura digitale. La constatata convergenza delle competenze e la promozione di iniziative collaborative e coordinate si presentano così come la via obbligata per la rivendicazione del ruolo dell'informatica umanistica nelle attività pratiche e di ricerca legate alla gestione dell'intero lascito della nostra tradizione culturale.

La mancanza di una base istituzionale definita impone anche un altro compito prioritario che deve impegnare in modo costante e imperativo l'attività dell'Associazione. Come assicurare la formazione e valutare le competenze degli studiosi operanti nel campo dell'informatica umanistica? Mancando un esplicito riconoscimento della specificità disciplinare di questo preciso settore di ricerca, la valutazione dell'adeguatezza scientifica delle indagini condotte in questo campo solleva problemi affatto particolari in un ambito già di per sé problematico e irto di difficoltà irrisolte. Quali le comunità giudicanti? quali i criteri specifici di adeguatezza scientifico-disciplinare? quale il giudizio sulla pertinenza degli indici di valutazione bibliometrica? L'insieme di queste questioni ha motivato l'organizzazione della seconda sessione dei lavori del convegno, la cui

attenzione è stata rivolta a una ricognizione delle pratiche correnti e dei principi ispiratori di una valutazione adeguata e corretta della produzione scientifica rilevante per la ricerca in questo campo. Tale produzione pone immediatamente il problema, assolutamente inedito nei settori tradizionali di studio, del riconoscimento e della certificazione dei risultati ottenuti direttamente in forma digitale. Siamo ancora molto lontani dall'accettazione piena di questa forma di produzione scientifica e questa circostanza motiva in modo impellente l'attenzione e l'impegno dell'Associazione per l'insieme di tali questioni. A testimoniare il dibattito di questa sessione pubblichiamo qui le riflessioni di Tito Orlandi circa la natura dell'Informatica Umanistica in quanto disciplina metodologica autonoma, radicata nei metodi formali, e l'intervento di Giovanni Solimine e Chiara Faggiolani, dedicato invece a una disanima critica dei metodi di valutazione della ricerca umanistica, nella dialettica tra peer-review e bibliometria.

La terza sessione del convegno è stata infine dedicata ad una ricognizione delle esperienze e attività in corso nel campo disciplinare che l'Associazione intende rappresentare. All'appello che abbiamo avanzato alla comunità di studi e di pratiche che si riconosce sotto la categoria delle Digital Humanities hanno risposto in molti e dopo una selezione condotta secondo i criteri della peer review sono stati approvati gli interventi di cui in questo volume pubblichiamo le relazioni scritte. In linea generale possiamo individuare tre aree tematiche in cui articolare i lavori, sebbene, come ci si può aspettare, numerose siano le sovrapposizioni e interconnessioni:

1. architetture informative e interfacce degli archivi digitali;
2. esperienze, teorie e sperimentazioni di filologia digitale;
3. progetti e servizi innovativi per repository e biblioteche digitali.

Inaugura la prima sezione Pier Luigi Feliciati con *Convergere a valle. Lo studio del punto di vista degli utenti degli ambienti culturali digitali e l'esperienza del progetto "Una Città per gli Archivi"*. Il lavoro propone una interessante riflessione circa l'importanza della progettazione delle interfacce e delle modalità di fruizione nei sistemi di disseminazione di oggetti e contenuti culturali sul Web, richiamando l'utilità dell'adozione di metodologie progettuali basate sull'*user centered design*. Le osservazioni teoriche e metodologiche

vengono peraltro messe alla prova nel lavoro condotto per la realizzazione del progetto "Una città per gli archivi".

Su una linea di riflessione e un ambito di sperimentazione affini si pone l'intervento di Michele Mauri, Paolo Ciuccarelli *Il ruolo dell'Information Visualization nella progettazione di interfacce per archivi digitali eterogenei* che, a partire dagli studi sulle nuove modalità di *information visualization* e di design delle interfacce per l'accesso a collezioni digitali, illustra una applicazione di tali strumenti all'archivio personale dell'architetto Luciano Baldessari.

L'articolo di Francesca Mambelli *Una risorsa online per la storia dell'arte: il database della Fondazione Federico Zeri* descrive il progetto di digitalizzazione e catalogazione di uno tra i più rilevanti archivi fotografici dedicati alla storia dell'arte italiana, documentando e discutendo le scelte e le modalità di utilizzazione degli standard descrittivi (in particolare la scheda F) e delineando i possibili scenari di sviluppo per questa importantissima risorsa.

Si passa alla sezione sulla filologia digitale nell'interessantissimo lavoro di Gioele Barabucci, Angelo Di Iorio, Fabio Vitali, *Stemma codicum: analisi e generazione semi-automatica*, in cui viene presentato un metodo semi-automatico per la costruzione dello *stemma codicum* di un insieme di documenti con varianti, codificati in un formato XML. Si tratta di un problema che è stato oggetto di numerosi studi nell'ambito della filologia digitale e sul quale sono stati applicati diversi paradigmi computazionali e algoritmi. La soluzione proposta qui elabora alcuni algoritmi già noti per il calcolo automatico di differenze e similarità tra alberi XML affiancando a essi un approccio interattivo e controllato dallo studioso. Questi è infatti in grado di definire in modo dichiarativo i parametri e criteri in base ai quali raffinare il confronto e la generazione dello stemma (selezione degli errori più significativi, definizione di nuove classi di errori, integrazione con informazioni non testuali).

Ancora in ambito filologico si pone *Il progetto Visionary Cross: verso un'edizione digitale multimediale e distribuita* di Roberto Rosselli Del Turco. Oggetto del lavoro è l'edizione digitale di una complessa tradizione di testi e iscrizioni monumentali di era anglosassone che vertono sul "culto della Croce" e, in particolare, sulla *Leggenda della Vera Croce*, assai diffusa in epoca medievale. Oltre all'interesse della tradizione in sé, questo progetto si segnala per il tentativo di ripensare

teoricamente il concetto di edizione critica digitale, anche in virtù della necessità di rendere conto delle complesse relazioni tra oggetti testuali e documentali così diversi come testi manoscritti e monumenti istoriati, che richiedono anche la sperimentazione di tecnologie di visualizzazione 3D.

Anche il lavoro di Paolo Monella *Many witnesses, many layers: the digital scholarly edition of the Iudicium coci et pistoris* (Anth. Lat. 199 Riese) si concentra su un'edizione scientifica digitale di un testo con una tradizione complessa. Basandosi sulle riflessioni di due studiosi che hanno dato un grandissimo contributo alla filologia informatica in Italia, Tito Orlandi e Raul Mordenti, Monella propone un processo ecdotico in cui ogni testimone è soggetto a un lavoro di codifica a più livelli (grafico, alfabetico e linguistico) e la collazione viene condotta in modo parallelo su ciascuno di tali livelli. In particolare si segnala la rivalutazione del ruolo del sistema grafemico e alfabetico nel confronto tra i documenti testuali.

Gli ultimi lavori pubblicati in questo volume sono invece relativi alla creazione e all'accrescimenti di servizi e funzioni di grandi repository di risorse digitali. Inizia in modo forse eccentrico la sezione il lavoro di Caterina Bernardini. Nel suo *Immaginando il futuro digitale della traduzione letteraria: la traduzione dell'edizione 1860 di Leaves of Grass per il Walt Whitman Archive* propone una riflessione sulle possibilità innovative che si aprono per la traduzione letteraria grazie alla disponibilità dei materiali pubblicati nei grandi archivi digitali d'autore come il *Whitman Archive* – progettati secondo il modello del *Rossetti Archive* realizzato da Jerome McGann. Le nuove modalità di pubblicazione digitale, come già avviene per le edizioni scientifiche dei testi, permettono infatti di rendere conto della pluralità e apertura che caratterizzano il processo di traduzione.

Rimaniamo nel dominio della traduzione con il lavoro di Federico Boschetti *La localizzazione del Perseus Project in lingua italiana* che illustra il progetto – in corso presso l'Istituto di Linguistica Computazionale del CNR di Pisa – per affiancare ai testi dei classici greci e latini contenuti nel *Perseus Project* le rispettive traduzioni italiane già disponibili in formato digitale o da digitalizzare. Il progetto oltre a fornire un complesso sistema di allineamento parallelo dei testi, si propone di fornire agli studiosi un vero e proprio ambiente di lavoro per l'analisi e l'annotazione degli stessi.

Maria Rosaria Falcone nel suo *Il progetto ENArC. Attività didattiche innovative e creazione di archivi digitali* illustra le attività svolte nell'ambito del progetto internazionale ENArC (European Network on Archival Cooperation - EU, Culture Programme 2007-2013) che ha come finalità la creazione di una rete europea per la tutela, la conservazione e la fruibilità della documentazione della cultura scritta. In particolare l'intervento si sofferma sull'uso di questa risorse per la didattica della Paleografia e della Diplomatica con l'impiego delle nuove tecnologie e sul portale Monasterium.NET che rende accessibili circa 230.000 documenti e 500.000 riproduzioni di documenti conservati in Italia e in Spagna, tutti compiutamente descritti mediante il software MOM-Ca, sviluppato appositamente per la gestione della documentazione medievale.

Maurizio Lana con *Letteratura latina e ebook. Una prospettiva della biblioteca digitale digilibLT* descrive finalità architettura e servizi di questo progetto dedicato alle pubblicazione in rete di opere del latino tardo pagano. Da segnalare la scelta consapevole e strategica di adottare una politica di distribuzione *open access* dei testi digitalizzati, attraverso licenze *Creative Commons*, e quella di distribuire i testi in vari formati, non ultimo quello *ePub*, lo standard open per il mondo *e-book*. In questo modo non solo gli studiosi di settore ma anche il più vasto pubblico dei potenziali fruitori è in grado di accedere e di consultare i testi in modi e contesti di uso altamente personalizzati.

Chiude il volume l'articolo di Marco Giunti, Giuseppe Sergioli e Giuliano Vivanet *L'ontologia BIA-Net: una base per la ricerca di informazioni secondo rapporti di rilevanza nella Bibliotheca Iuris Antiqui*. Il lavoro presenta una interessante evoluzione del prestigioso progetto dedicato al corpus legislativo dell'antichità e della sua interfaccia Web BIA-Net. Si tratta in particolare di una ontologia formale (Ontologia BIA-Net), che descrive le classi concettuali e relazioni interne e reciproche dei tre archivi che compongono il progetto (archivio bibliografico, thesaurus e fonti). In chiusura del lavoro vengono anticipati gli sviluppi previsti verso la creazioni di due applicazioni di tipo semantico, un motore di ricerca e un browser, tramite cui rendere accessibili i dati dell'archivio.

I lavori del convegno e le relazioni presentate hanno fatto emergere un quadro assai vitale di iniziative e di risultati, com'è manifestamente testimoniato dai contributi qui raccolti. Il panorama attuale delle

Digital Humanities o, come preferiamo, dell'Informatica Umanistica italiana, conferma dunque l'importanza di una tradizione di ricerca e di riflessione che risale almeno agli anni 60, e che ha prodotto lavori e contributi di grande rilevanza. Dobbiamo ammettere che tale tradizione scientifica, per motivi prima di tutto linguistici ma anche culturali, ha avuto una circolazione e una influenza internazionale molto inferiore al suo valore intrinseco, e un riconoscimento accademico nazionale se possibile ancora più limitato. Crediamo che la fondazione di una *Associazione per l'Informatica Umanistica e la Cultura Digitale* e il successo delle sue iniziative scientifiche e istituzionali sarà presupposto e stimolo perché gli umanisti digitali italiani abbiano infine il giusto riconoscimento.

Bibliografia

- ROSNER, D., ROCCETTI, M., and MARFIA, G., *Transitioning Tradition? Examining the Digitization of Cultural Practices*, «Communications of the ACM», (forthcoming).
- MCGANN, J. J., *Memory Now*, «4Humanities», 19 August 2012, <http://4humanities.org/2012/08/jerome-j-mcgann-memory-now-2/> .

PARTE I

DIGITAL HUMANITIES: VALUTAZIONE E INFRASTRUTTURE

1. La ricerca, la valutazione e la diffusione dei risultati nell'informatica umanistica. Problematiche aperte

*Tito Orlandi*¹

Il titolo di questa sessione mi sembra particolarmente felice, in quanto mette in evidenza la ricerca. È sotto gli occhi di tutti, per la verità, che nella maggior parte dei casi l'informatica, o sarebbe meglio dire i computer, viene presa in considerazione dagli studiosi di discipline umanistiche nei suoi aspetti più attraentemente applicativi: la navigazione in rete, la disponibilità di documenti per una consultazione facile e immediata, la costituzione di raccolte di testi, immagini, composizioni sonore, di ogni epoca, i contatti fra le persone resi immediati e flessibili, senza scendere ad altri più banali. Ma è bene tenere questi aspetti distinti dalla ricerca nelle discipline umanistiche (da qui in avanti ricerca, *tout court*) che si sforza di vedere nell'informatica un nuovo modo di considerare i problemi delle discipline umanistiche, rigoroso e formalmente corretto, estremamente analitico, e dunque riferito alle metodologie oltre e meglio che al puro aspetto di conservazione e di comunicazione di documenti e di contributi critici.

Di questa attività scientifica e della sua valutazione intendo trattare sinteticamente in questo intervento. Ma vorrei sbarazzarmi subito di una delle due possibilità di impostare questa materia, quella che intende sottoporre la valutazione a criteri quantitativi, di cui del resto altri parlerà in seguito. La ritengo una scelta inopportuna, e comunque la conoscenza approfondita di una tale procedura non rientra nelle mie competenze.

Rimanendo allora nel campo di una valutazione qualitativa, ritengo importante la distinzione fra due possibili tipi di ricerca,

¹ Accademia dei Lincei. e-mail: orlandi@rmcisadu.let.uniroma1.it.

ciascuno dei quali richiede sue proprie e diverse analisi di pertinenza e di efficacia. Il primo si rivolge a quello che definirei aspetto sociologico, e cioè al cambiamento di mentalità spesso inconsapevole cui sono soggetti gli studiosi in presenza di uno strumento così incredibilmente potente e pervasivo, la cui esistenza non poteva essere fino a tempi recenti neppure concepita. È un campo affascinante, ma mi sembra ausiliare e sostanzialmente estraneo a quanto qui vogliamo discutere. Lo menziono perché spesso accade che studi ad esso dedicati facciano passare il loro oggetto per il risultato di una ricerca che in realtà non c'è stata. Intendo dire che condurre un'analisi stilistica sulla base della statistica dell'uso di determinate parole o sintagmi, per fare un esempio, è certo indice di cambiamento di mentalità, ma non è necessariamente prova di una ricerca effettuata. Siamo invece in presenza di vera ricerca quando lo studioso di discipline umanistiche vede nel computer la concezione di base in virtù della quale esso funziona e fornisce dei risultati, e cioè la Macchina di Turing come automazione di procedimenti logici, con quanto ne consegue in linea teorica, e cioè la scienza degli algoritmi, la computazione, la ricorsività, i linguaggi artificiali. Ma tale visione implica anche, e questo viene spesso trascurato, fare seriamente i conti con le teorie proprie di discipline quali la semiotica, la scienza dei modelli, e la scienza dei sistemi.

Infatti la Macchina di Turing ci indica che ciò che viene sottoposto a trattamento automatico sono dei simboli, non il loro significato. L'essenza dei simboli e il loro rapporto con il significato sono il tipico oggetto della semiotica, che qui si trasforma da scienza sui cui esiti si può discutere all'infinito, e in certo senso anche trascurabile, in terreno di scelte delicate e imprescindibili, senza le quali la ricerca rimane priva di fondamento. Basterà ricordare il caso della codifica, che tuttavia non è il solo. Ancora, si riflette raramente sul fatto che il computer dà la possibilità di fornire una rappresentazione *in movimento* della realtà studiata, per la prima volta nella storia degli studi. Qui appunto interviene la teoria dei modelli, che guida questo tipo di realizzazioni. Un passo ulteriore è fornito dai sistemi, che possono far interagire più modelli evidenziando i reciproci rapporti che essi possono avere, e la famosa constatazione che in un sistema il tutto è qualcosa di più della somma delle sue parti. In realtà di fronte a queste teorie l'informatica in quanto tale finisce per annullarsi,

creando, se vogliamo, un problema di tassonomia delle ricerche. Ma la funzione del computer, macchina algoritmica prodotto esclusivo dell'informatica, rimane essenziale perché esso solo fornisce le *costrizioni* oggettive in virtù delle quali un programma agisce o non agisce. Questo tipo di obiettività è molto importante nella valutazione della ricerca.

Solo in questo modo diventa possibile attuare una valutazione che distingua le pure applicazioni tecnologiche, che in sostanza usano il computer e la rete come un mezzo non modificabile (il contrario di quello per cui è stato concepito il *programma*), dalle applicazioni di carattere metodologico. Concepire un programma infatti significa andare alla base delle metodologie che dirigono una ricerca. Sfruttare un programma preconfezionato significa rinunciare a quanto di originale si cerca di proporre. È vero che in questo modo si ridimensiona il valore della maggior parte delle imprese esistenti, ma solo per quanto riguarda il punto di vista dell'informatica umanistica. Nei diversi campi applicativi esse mantengono pienamente la loro validità.

Nel campo specifico dell'informatica umanistica è essenziale riconoscere il *valore aggiunto* che procura alla ricerca il connubio fra l'informatica, intesa in senso teorico, e le discipline umanistiche, qui distinte dalle scienze umane di cui non intendo occuparmi. Se questo valore aggiunto va solo alle discipline umanistiche, non si può pretendere di fare ricerca in informatica umanistica. Benemerite istituzioni come il Progetto Gutenberg o LiberLiber o il Thesaurus Linguae Graecae, e infiniti Musei e Archivi on line contribuiscono senza dubbio alla ricerca umanistica, ma non possono considerarsi ricerca in informatica umanistica.

Si deve identificare piuttosto una svolta metodologica, per la quale consolidate vedute devono mutare, se si vogliono usare correttamente strumenti informatici all'analisi di determinati soggetti. L'esempio più chiaro che possa essere portato in questo caso è quello della linguistica, settore nel quale si sono determinati cambiamenti sia nel campo umanistico sia in quello informatico. È noto che agli inizi i tentativi di traduzione automatica si basarono su una concezione ingenua della linguistica e del testo; concezione che del resto era quella corrente fino allora. Gli esperimenti, pur infruttuosi, portarono ad una rivoluzione nelle teorie delle parti del discorso e delle strutture grammaticali,

senza contare la presa d'atto di una corrispondenza formale fra espressioni in diverse lingue, per cui si ricorre oggi all'uso di *corpora* paralleli. Oltre a questo, è interessante registrare che proprio le concezioni linguistiche di Chomsky portarono a formulare i linguaggi evoluti in informatica.

Per concludere, desideriamo ribadire cose che sono state illustrate molte volte, da noi e da altri, ma che sono sempre attuali dal momento che non hanno trovato attuazione da chi ne aveva l'autorità. Prima di tutto, per avere una valutazione corretta, occorre che si riconosca la realtà della disciplina individuale e individuata *informatica umanistica*, perché solo all'interno di essa si possono puntualizzare i criteri relativi alle ricerche che abbiamo cercato di delineare sopra. In secondo luogo tale riconoscimento deve avere degli esiti formalizzati a livello accademico, altrimenti non vi sarà mai una istituzione che prepari gli esperti di informatica umanistica da immettere nelle diverse discipline. Come abbiamo visto, non basta affiancare esperti informatici a ricercatori umanistici, perché sono in gioco rapporti metodologici che richiedono una competenza bivalente. Nei paesi anglosassoni questo problema si va risolvendo con la prassi accademica, come è usuale in quell'ambiente; da noi occorre continuare a chiedere dei provvedimenti legislativi o regolamentari che allo stato attuale non possono provenire dalle singole Università. Un ruolo in questo senso può svolgere l'Accademia, con l'approfondire le tematiche dell'informatica umanistica, e mettere sotto gli occhi di chi abbia la giusta attenzione i problemi che abbiamo delineato. Per questo il Centro Linceo Interdisciplinare presso l'Accademia dei Lincei, che al momento ho l'onore di dirigere, ha scelto come una delle sue linee di ricerca proprio l'informatica umanistica, con l'auspicio che presto o tardi il settore entri a far parte in modo giusto del mondo delle Facoltà di lettere e filosofia.

2. La valutazione della ricerca umanistica: tra peer-review e bibliometria

Chiara Faggiolani¹, Giovanni Solimine²

Abstract. L'intervento, preparato congiuntamente ma sviluppato separatamente dai due autori (Giovanni Solimine: paragrafi 1 e 3; Chiara Faggiolani: paragrafi 2, 4 e 5) si apre con una riflessione a carattere introduttivo su ciò che sta accadendo nel campo della valutazione della ricerca in Italia, anche sulla spinta delle attività promosse dall'ANVUR.

Passa poi ad esaminare i problemi specifici legati alla valutazione nel delle discipline umanistiche e formula alcune proposte sul ruolo che le biblioteche, la bibliografia e la biblioteconomia potrebbero esercitare per una diffusione e una più rigorosa applicazione degli studi bibliometrici, promuovendo ad esempio un maggiore utilizzo di alcuni strumenti direttamente legati alle biblioteche, come la *Library Catalog Analysis* (LCA). L'intervento prosegue discutendo alcune questioni specificamente legate alla valutazione delle pubblicazioni digitali.

Nelle conclusioni si descrivono i possibili obiettivi di una ricerca nel settore della valutazione della ricerca umanistica e il percorso da compiere per poterli raggiungere.

Parole chiave: peer review, bibliometria, valutazione ricerca, VQR, ASN, web impact factor, usage factor, LCA, biblioteche, biblioteconomia.

¹ Dipartimento di Scienze documentarie, linguistico-filologiche e geografiche, Università di Roma Sapienza. e-mail: chiara.faggiolani@uniroma1.it.

² Dipartimento di Scienze documentarie, linguistico-filologiche e geografiche, Università di Roma Sapienza. e-mail: giovanni.solimine@uniroma1.it.

2.1. Cosa sta accadendo nel campo della valutazione in Italia

Il tema della valutazione della ricerca è da qualche tempo al centro del dibattito nel mondo universitario, ma non solo: il definanziamento del sistema universitario pubblico posto in essere dai governi che si sono succeduti negli ultimi anni, il declino della ricerca scientifica in Italia, i temi legati alla crisi economica e al percorso attraverso il quale costruire un futuro di crescita hanno contribuito a porre la questione all'attenzione di una più vasta platea di osservatori e commentatori. Noi stessi ci siamo già occupati recentemente di tali questioni (Faggiolani, Solimine, 2012).

La questione della valutazione della ricerca si collega, ovviamente, al più ampio tema della *accountability* dell'università italiana, che da circa un ventennio si è imposto come naturale conseguenza delle leggi che regolamentano l'autonomia normativa, organizzativa, finanziaria e contabile, didattica e scientifica dell'università italiana³.

È ormai ampiamente condivisa l'esigenza di individuare criteri che possano oggettivamente misurare dal punto di vista quantitativo e analizzare dal punto di vista qualitativo la produttività scientifica di docenti e ricercatori e delle strutture (università e dipartimenti) in cui essi operano, rilevando il 'valore' e l'impatto di ciò che la ricerca universitaria produce.

A tale scopo è stata istituita l'Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR)⁴, che attualmente sta gestendo l'esercizio di Valutazione della qualità della ricerca (VQR) per il periodo 2004-2010, che coinvolge oltre 68.000 docenti e ricercatori delle università e degli Enti di ricerca: i risultati di questo articolato processo, che mobilerà circa diecimila valutatori, avranno effetto a livello nazionale sulla distribuzione del fondo di finanziamento ordinario. Metodi e criteri adottati in queste circostanze sono stati ampiamente criticati da più parti⁵.

Per inciso, potremmo aggiungere anche che un tale meccanismo

³ A partire dalla L. 168/1989 fino alla recente L. 240/2010, la cosiddetta "riforma Gelmini".

⁴ Cfr. <http://www.anvur.org/>.

⁵ Tali critiche e la discussione che le accompagna sono ampiamente documentate sul sito ROARS (Return On Academic Research) <http://www.roars.it/>.

sarebbe ancor più giustificato se il Governo mettesse a disposizione del mondo della ricerca e della formazione universitaria risorse significative, altrimenti rischieremmo di fare tanto rumore per nulla. Ma questa è un'altra questione che al momento in cui scriviamo (gennaio 2013) è avvolta in una nebbia fittissima e che forse potrà essere affrontata solo all'indomani delle elezioni politiche del 24-25 febbraio 2013, quando il nuovo parlamento e il nuovo governo dovranno decidere se puntare sull'istruzione e sulla ricerca per uscire dall'attuale crisi economica e se avvertiranno il dovere di accorciare, quanto meno, il divario che ci separa da altri paesi occidentali ed europei, facendo incamminare anche l'Italia sulla strada tracciata dalla Commissione Europea, che col programma Horizon 2020⁶ ha individuato tre motori di crescita per preparare l'economia dell'UE ad affrontare le sfide del decennio che stiamo vivendo: crescita intelligente (promuovendo la conoscenza, l'innovazione, l'istruzione e la società digitale), crescita sostenibile (rendendo la nostra produzione più efficiente sotto il profilo dell'uso delle risorse, rilanciando nel contempo la nostra competitività) e crescita inclusiva (incentivando la partecipazione al mercato del lavoro, l'acquisizione di competenze e la lotta alla povertà).

Si tratta, come si può vedere, di questioni piuttosto complesse e di ampia portata. Un problema nel problema è costituito dalla individuazione di criteri da applicare alla valutazione della ricerca umanistica.

2.2. Problemi specifici legati alla valutazione della ricerca umanistica

Particolarmente delicato è il tema della valutazione della ricerca nell'ambito delle scienze umane e sociali. Infatti, mentre nella comunità degli studiosi di ambito scientifico-tecnologico sono universalmente accettati anche a livello internazionale – se pure con qualche eccezione (Figà Talamanca 2000) – i metodi dell'analisi citazionale e alcuni indicatori bibliometrici (Impact Factor, H-Index etc.), che misurano in termini quantitativi l'impatto della produzione scientifica, lo stesso non si può dire per le scienze umane e sociali.

La questione preliminare da affrontare a nostro avviso non è quella metodologica ma quella dell'oggetto della valutazione: una volta

⁶ Cfr. http://ec.europa.eu/research/horizon2020/index_en.cfm.

delineato, infatti, la definizione del metodo da utilizzare sarà piuttosto naturale.

A tale proposito è utile partire dalle differenze tra scienze dure e scienze umane (cfr. Tab. 1) per dimostrare per quale ragione gli strumenti bibliometrici, che per le prime sembrano così efficaci, per le seconde – almeno in questo momento – non possono essere utilizzati con successo.

	SCIENZE DURE	SCIENZE UMANE
Oggetto	<i>Fenomeno naturale</i>	<i>Fenomeno prodotto dalla mente umana</i>
Relazione tra oggetto e ricercatore	<i>Osservatori interscambiabili</i>	<i>Esperienza personale che coinvolge l'individuo nella sua interezza</i>
Prospettiva	<i>Regolarità nei modelli dei dati, leggi scientifiche</i>	<i>Aspetti unici e irriducibili</i>
Linguaggio	<i>Linguaggio matematico</i>	<i>Linguaggio naturale</i>
Organizzazione	<i>Ricerca internazionale</i>	<i>Confini sfumati tra comunità scientifica e società (dibattito pubblico)</i>
Crescita di conoscenza	<i>Incrementale</i>	<i>Oggetto percepito come un intero</i>
Unità di base	<i>Gruppo di ricerca</i>	<i>Individuale</i>
Attività di ricerca	<i>Progetti a breve termine</i>	<i>Investimento personale nel lavoro di tutta una vita</i>
Velocità di circolazione delle idee	<i>Alta</i>	<i>Bassa</i>
Tipologia di pubblicazione	<i>Articoli di periodici</i>	<i>Libri</i>
Lingua di pubblicazione	<i>Inglese</i>	<i>Lingua nazionale</i>
Livello di aggregazione	<i>Gruppo di ricerca</i>	<i>Individuale</i>
Orizzonte temporale	<i>Breve (2 generazioni di PHD)</i>	<i>Lungo tutta una vita</i>

Tab. 2.1. Comparazione tra scienze dure e scienze umane (Moed 2008).

La ricerca in ambito umanistico è strutturalmente diversa da quella prodotta nelle scienze dure, di seguito si evidenziano le principali differenze che impattano sulla questione metodologica:

1. **Tipologia di pubblicazione:** l'articolo su rivista – sul quale tradizionalmente si basano gli indicatori di impatto – non è la forma di pubblicazione più diffusa tra gli umanisti, la monografia ha un ruolo preminente per ragioni senza dubbio riconducibili alla natura cumulativa del sapere umanistico. Questo si serve principalmente dell'opera monografica come strumento comunicativo denso, complesso, enciclopedico rispetto all'argomento trattato. Per un ricercatore che opera nell'ambito delle scienze umane l'articolo è, dunque, solo la premessa per una riflessione di più ampia portata, suscettibile di essere sviluppata in ricerche e opere di maggiore respiro. Le scienze umane sono caratterizzate, di fatto, da "quattro letterature" (Hicks 2004, pp. 473-496):

- la rivista scientifica;
- la monografia, che è la forma di pubblicazione più diffusa;
- la letteratura nazionale;
- la produzione non scientifica (editoriali etc.)⁷.

2. **Comportamento citazionale:** in ambito umanistico non è sufficiente contabilizzare le citazioni, perché non sempre esse corrispondono ad un giudizio di merito di segno positivo. Di fatto l'IF non distingue le citazioni bibliografiche e le autocitazioni, le citazioni positive da quelle negative. Un articolo molto citato potrebbe esserlo come esempio di cattiva ricerca. Sebbene possa sembrare un meccanismo semplice e lineare, in realtà il comportamento citazionale può essere condizionato da variabili anche di carattere "sociale" non sempre riconducibili al piano strettamente scientifico. Ad una "teoria normativa" (Merton 1988), che afferma che il conteggio delle citazioni rende conto quasi naturalmente dell'impatto e del prestigio di un lavoro scientifico, si contrappone una "teoria socio-costruttivista" che, al contrario, suggerisce che il comportamento citazionale è condizionato da

⁷ È interessante notare come nel REF 2014 (<http://www.ref.ac.uk/>) l'equivalente britannico della VQR italiana, una parte importante e molto costosa dell'attività di valutazione consisterà proprio nella valutazione dei lavori non scientifici, di cui si vuole valutare l'impatto. L'agenzia di valutazione inglese non si propone di eliminare prodotti della ricerca ma di capire gli effetti del lavoro non scientifico svolto dai ricercatori: basti pensare alla diffusione di regole e informazioni rivolta ai professionisti, agli articoli non scientifici che appaiono sui quotidiani.

molteplici fattori che ne possono manipolare fortemente il significato (Gilbert 1977). Questa seconda interpretazione è particolarmente significativa in ambito umanistico poiché – come anticipato – se nelle scienze dure le citazioni possono dirsi piuttosto asettiche, in campo umanistico la citazione può essere anche "critica" o "problematica", ovvero può fare riferimento al lavoro citato non solo per approvarlo e sostenerlo ma anche per confutarlo o valutarlo negativamente.

3. **Co-autoraggio:** molto meno frequente è, in ambito umanistico, l'abitudine a pubblicare scritti in collaborazione fra più autori, pratica che fa aumentare l'impatto delle pubblicazioni stesse poiché, come è noto, le opere in co-autoraggio vengono citate di più.
4. **Lingua di pubblicazione:** molto spesso la lingua privilegiata nel campo umanistico e sociale è la lingua madre: ciò dipende dalle tematiche di carattere locale, oggetto di studio di queste discipline, maggiormente radicate nel loro contesto di quanto non accada nelle scienze naturali. Gli studiosi di scienze umane e sociali sono molto influenzati nelle loro scelte di ricerca da quanto accade a livello nazionale, da quelli che sono gli argomenti prevalenti in ambito pubblico e ciò influisce anche sulla lingua di pubblicazione. L'uso della lingua madre rispetto alla comunità scientifica di riferimento – locale appunto – dà garanzia di migliore comprensione e impatto sulla propria comunità scientifica (Hicks 2004, pp. 473-496).
5. **Prospettiva storica delle scienze umane:** il valore scientifico di un lavoro di area umanistica è prolungato negli anni e il suo impatto è misurabile solo su un lungo periodo di tempo. Se nelle scienze naturali l'avanzamento della ricerca procede con il superamento delle ricerche precedenti, nelle scienze umane la ricerca procede per progressiva accumulazione e approfondimento. È per questa ragione che per gli indicatori bibliometrici tradizionali, che utilizzano un *range* temporale piuttosto stretto che va dai due ai cinque anni (come l'IF o il 5-year IF) si può parlare di mancanza di prospettiva storica che, in ultima analisi, è il tratto peculiare di ogni riflessione di carattere umanistico.

Sulla base di queste considerazioni, c'è chi rifiuta qualsiasi prospettiva di valutazione bibliometrica per le scienze storiche, filosofiche, filologico-letterarie, giuridiche, economiche, politiche e sociali. Va invece assolutamente evitato che col pretesto che la valutazione in area umanistica è molto difficile si finisca col sostenere che essa è impossibile.

Si ritiene indispensabile operare in direzione di una coerenza dei criteri di valutazione anche rispetto alle scienze dure, riconoscendo però le specificità della ricerca nelle scienze umane. È necessario, cioè, mettere a punto un sistema di valutazione che sia rispettoso delle specificità dei diversi saperi e delle loro modalità di disseminazione. Forzare la ricerca umanistica verso modelli propri di ambiti disciplinari differenti potrebbe essere pericoloso, pena, nel tempo, un peggioramento della ricerca dovuto ad un cambiamento delle strategie.

Non è, infatti, difficile immaginare che invece di pensare a pubblicare buona ricerca i ricercatori inizierebbero a preoccuparsi di pubblicare ricerca meno buona ma su riviste con elevato *impact factor*; così come potrebbe esserci il rischio di un appiattimento su tematiche di ricerca *mainstream* poichè più citate, a scapito di quella pluralità tematica e metodologica che contraddistingue da sempre le discipline umanistiche.

2.3. Il ruolo delle biblioteche, della bibliografia e della biblioteconomia

La grande assente dalle applicazioni della Bibliometria è proprio la disciplina da cui essa in qualche modo dovrebbe derivare, vale a dire la Bibliografia – intesa come disciplina che si occupa di tutto ciò che riguarda l'informazione bibliografica, ivi comprese le pratiche citazionali –, oppure la Biblioteconomia, se vogliamo riferirci agli studi sulla circolazione dei documenti e quindi anche a quella parte della comunicazione scientifica che passa attraverso i servizi pubblici di organizzazione dei documenti in funzioni dell'uso, vale a dire le biblioteche.

Siamo convinti che un legame più stretto della Bibliometria con la Bibliografia e la Biblioteconomia, e quindi anche con le biblioteche, sarebbe di grande utilità.

Parimenti, siamo convinti anche che una maggiore attenzione dei

bibliografi e dei biblioteconomi agli studi di analisi bibliometrica potrebbe fornire un supporto specifico alla elaborazione di criteri di valutazione rigorosi e affidabili, migliorando così la qualità della prassi valutativa della ricerca scientifica in generale e soprattutto nel caso delle scienze umane che presentano non pochi tratti peculiari.

Specie per quanto riguarda l'Italia il cammino da percorrere è molto lungo. A differenza degli Stati Uniti e di altri paesi europei, tra i quali si segnala la Spagna, non vi è nel nostro Paese una tradizione di studi bibliometrici, e comunque non ad opera di bibliotecari e biblioteconomi: se escludiamo qualche contributo offerto anni fa da Alfredo Serrai (1984, pp. 199-229), dobbiamo attendere questi ultimi anni per trovare qualche segnale di interesse verso l'argomento (Cassella 2010; Cassella, Bozzarelli 2011, pp. 66,78; Comba 2003, p. 65-76; Tammaro 2001, pp. 104-107; Guerrini 2009, pp. 7-17; De Bellis 2009) anche sull'onda dello sviluppo dell'editoria digitale e del movimento open access. Mancano anche dati di spogli di periodici e di citazioni, né esistono indagini empiriche o studi approfonditi sulla copertura delle collezioni bibliografiche, dei cataloghi e degli OPAC riguardo alla produzione editoriale in determinati settori disciplinari o da parte di determinati studiosi.

Le biblioteche hanno un ruolo importante da giocare nella valutazione della ricerca in ambito umanistico: non va dimenticato, infatti, che le pubblicazioni monografiche costituiscono il principale strumento attraverso il quale vengono comunicati i risultati della ricerca in questo settore e che esse sono di conseguenza anche la principale fonte delle citazioni che figurano negli scritti degli studiosi di ambito umanistico. Inoltre, proprio perché la valutazione della ricerca umanistica ha bisogno di fonti che coprano un arco temporale molto ampio, le biblioteche costituiscono anche per questo motivo il naturale laboratorio per l'analisi bibliometrica nelle scienze umane e sociali, pur nella consapevolezza che la bibliometria da sola non copre tutte le esigenze.

Non si può affermare, quindi, che non esistano le banche dati di riferimento per chi volesse applicare la bibliometria anche nel nostro settore: per tale finalità possono essere utilizzati i cataloghi delle biblioteche.

Per la valutazione dell'impatto delle monografie alcuni studiosi propongono di adottare la LCA (*Library Catalog Analysis*), rilevando se

una determinata pubblicazione è presente (e in quante copie) all'interno di un insieme selezionato di prestigiose biblioteche.

La LCA è definita come «l'applicazione di tecniche informetriche e bibliometriche a un insieme di cataloghi di biblioteche ed è focalizzata sul suo valore come strumento nello studio delle scienze umane e sociali. Questo strumento propone un modello analogo alla tradizionale analisi citazionale effettuata per gli articoli dei periodici, ma applicata agli OPAC per quanto riguarda i volumi, e illustra come la tecnica di mappatura tematica possa essere messa a frutto quale potente strumento per la valutazione delle monografie come produzioni intellettuali di ricerca a livello di singolo ricercatore, di dipartimento o come intera produzione di un paese o di un editore» (Torres Salinas, Moed, 2009).

Uno specifico contributo degli studi sulla valutazione delle biblioteche e delle loro raccolte può consistere nell'individuare gli istituti più autorevoli da assumere come riferimento per una determinata area disciplinare ed analizzare le politiche di acquisizione che essi praticano: la LCA, infatti, può essere considerata uno strumento affidabile se le biblioteche considerate rispettano rigorosi e trasparenti criteri di selezione e sviluppo delle raccolte, verificabili attraverso una 'carta delle collezioni' resa pubblica, se esse adottano un filtro nell'accettazione dei doni, se studiosi qualificati esprimono elevati livelli di soddisfazione riguardo al livello di copertura bibliografica che tali biblioteche garantiscono, e così via.

Ciò che per il contesto italiano può sembrare una novità o un azzardo, non è tale a livello internazionale. Un documento che il maggiore network di biblioteche a livello internazionale ha dedicato al tema di cui qui si sta discutendo (*Research Assessment and the Role of the Library*)⁸ definisce in questo modo il ruolo delle biblioteche:

1. Per le istituzioni le biblioteche devono essere le fonti della conoscenza delle norme adottate per la valutazione della ricerca nelle diverse pratiche disciplinari;
2. le biblioteche devono sostenere un ambiente in cui le discipline possano svilupparsi e co-esistere con i vincoli di natura politica;

⁸ <http://www.oclc.org/resources/research/publications/library/2010/2010-01.pdf>.

3. le biblioteche devono gestire i dati sui risultati delle ricerche scientifiche su scala nazionale ed internazionale. Le biblioteche possono sostenere una gestione più efficiente dei dati sulla valutazione della ricerca;
4. le biblioteche devono assumersi la responsabilità di gestire e mantenere i depositi istituzionali a sostegno della ricerca;
5. le biblioteche devono fornire expertise nel campo della bibliometria per sostenere la valutazione della ricerca con una corretta interpretazione dei differenti indici bibliometrici;
6. le biblioteche devono produrre dati sull'uso delle risorse, contribuendo a combinare metriche basate sull'analisi citazionale a metriche basate sul numero dei download;
7. le biblioteche devono ribadire il loro ruolo attivo nella valutazione della ricerca. Devono assicurarsi di avere una parte nella pianificazione istituzionale della valutazione della ricerca ed evitare di avere un ruolo unicamente reattivo.

Per le biblioteche italiane il cammino è ancora lungo, ma la strada è indicata con chiarezza.

2.4. Gli strumenti a disposizione: il ruolo del mondo digitale nella valutazione della ricerca

Prima di prendere in esame, se pure brevemente, la questione metodologica pare opportuno ricordare che almeno tre possono essere gli oggetti della valutazione (Baccini 2010)⁹:

Qualità (interna)

Si tratta del riconoscimento che la ricerca è stata ben condotta in riferimento ai canoni prevalenti in un certo momento del tempo: ad esempio originalità, rilevanza, rigore metodologico, chiarezza nell'esposizione etc.

È evidente come questo concetto di qualità si presti ad essere valutato dalla comunità scientifica di riferimento che conosce e condivide i canoni di qualità medesimi. Così la logica alla base del giudizio di

⁹ In questa sede non si prende in esame l'impatto sociale, ovvero la ricaduta degli esiti della ricerca scientifica nella società, che pure sarebbe interessante misurare.

qualità risulta del tutto interno alla disciplina. Non solo, questo giudizio di qualità oltre a variare nel tempo, poiché nel tempo possono variare i canoni di riferimento, presenta una forte componente di soggettività.

Impatto

Per impatto si intende il riconoscimento tributato ad un contributo scientifico dalla comunità dei pari attraverso le citazioni in un dato intervallo temporale. Un articolo molto citato è un articolo con un elevato impatto nella comunità scientifica. A questo proposito va notato che difficilmente un lavoro scientifico di nicchia o che tratta un argomento "di frontiera" può essere un articolo di particolare impatto, sebbene condotto secondo canoni di qualità. Questo accade perché l'impatto esula dalle caratteristiche interne del prodotto e dipende invece, ad esempio, dalla moda disciplinare del momento. Questo rischio è particolarmente sentito nelle scienze umane in cui la frammentarietà, intesa come consolidata separazione disciplinare, rappresenta una caratteristica così evidente da penalizzare quelle ricerche che possono essere ritenute multidisciplinari o interdisciplinari.

Importanza

Per importanza di un contributo scientifico si intende la capacità di influenzare nel lungo periodo le ricerche condotte da altri e di aprire la strada a nuove vie per lo sviluppo della scienza. L'importanza di un contributo può essere apprezzata e valutata solo nel lungo periodo, specialmente nel caso delle scienze umane per quanto detto in apertura. Venendo alla scelta degli strumenti, è evidente che per quanto riguarda la qualità l'unico strumento valido non può che essere la *peer review*: l'unico modo per dire se un articolo è valido o meno è che qualcuno (competente) lo legga.

Per quanto riguarda la misurazione dell'impatto, in linea di principio è condivisibile l'idea che non esistono discipline rispetto alle quali gli indicatori bibliometrici non siano applicabili: esistono piuttosto discipline rispetto alle quali al momento attuale, i dati non sono disponibili rendendo gli indicatori bibliometrici del tutto inaffidabili (Baccini 2010, p. 198). La qualità, il significato e l'utilizzabilità degli indicatori bibliometrici dipendono, infatti, in primo luogo dall'affidabilità degli archivi e da come sono costruiti.

Allo stato attuale i database bibliografici e citazionali esistenti non hanno informazioni sufficienti per applicare l'analisi bibliometrica alle

scienze umane come avviene nelle scienze dure¹⁰: WoS e Scopus coprono una piccolissima parte della letteratura rilevante per le scienze umane (soprattutto per quanto riguarda le monografie) e Google Scholar non si configura ancora come una fonte del tutto attendibile.

Si potrebbe dire che la bibliometria non è applicabile alle scienze umane perché non esistono database costruiti sulla base delle modalità di diffusione della conoscenza che caratterizzano queste discipline. Proprio a questo proposito va segnalato che esistono diversi studi su come potrebbe essere articolato un database bibliometrico *ad hoc* per le scienze umane (Hicks, Wang, 2010).

Cosciente dei limiti dell'analisi bibliometrica per la valutazione della ricerca nelle scienze umane, l'ANVUR in un documento dello scorso marzo redatto da Andrea Bonaccorsi¹¹ ha individuato una serie di priorità da sviluppare nei prossimi anni:

- Supporto alla candidatura delle riviste in lingua italiana per l'indicizzazione WoS e Scopus;
- Rating delle riviste non indicizzate in lingua italiana, con procedure comparabili con analoghe esperienze internazionali¹²;
- Pubblicazione di informazioni sulle procedure editoriali e di selezione dei manoscritti da parte di editori nazionali;
- Creazione di un archivio di metadati e di referenze tratte da monografie in lingua italiana e di riviste italiane disponibili in formato digitale;
- Ricerca e sperimentazione di indicatori non citazionali.

Proprio rispetto a quest ultimo punto il mondo digitale gioca un ruolo di primo piano. Va riconosciuto, infatti, che se l'analisi citazionale è la

¹⁰ I limiti dei database citazionali esistenti nel caso delle scienze umane e sociali sono ben noti e discussi da una ampia letteratura: Tarantino (2006), Archambault et al. (2006), Nederhof (2006), Hellqvist (2010).

¹¹ http://www.anvur.org/sites/anvur-miur/files/valutazione_aree_umanistiche_e_sociali.pdf.

¹² In assenza di indicatori bibliometrici e di archivi affidabili, ma con la stessa necessità delle scienze dure di affiancare alla peer review strumenti quantitativi, nell'ambito delle scienze umane si è pensato di utilizzare la classificazione delle riviste per "il principio secondo il quale il valore del contenuto corrisponde tendenzialmente al valore del contenitore" (Galimberti 2012).

metodologia quantitativa più diffusa nella valutazione della ricerca, essa non è da considerarsi esaustiva e grazie all'evoluzione delle tecnologie e alla crescita dei contenuti ad accesso aperto si stanno aprendo nuove prospettive.

Lo sviluppo dell'editoria elettronica consente di ottenere dati molto dettagliati sulla circolazione e sull'utilizzo della pubblicazione, scardinando la centralità dell'autore e ponendo l'attenzione sul lettore (Cassella, Bozzarelli 2011).

L'*Usage Factor*, definito come il download della versione digitale di un articolo, ne è un esempio. Si tratta di un indicatore bibliometrico quantitativo complementare all'*Impact Factor* che utilizza il download della versione digitale dell'articolo. Questo presenta una serie di vantaggi: non solo l'utilizzo può essere tracciato dal momento in cui l'articolo viene messo online (come preprint, ad esempio) ma è una metrica che può essere allargata anche ad altre tipologie di documenti (capitoli di monografie, audio, video, etc.) non limitandosi, quindi, a misurare l'impatto del solo articolo scientifico. Per quanto riguarda i limiti, oltre al rischio della manipolazione dei dati – che comunque riguarda anche le metriche citazionali tradizionali – ci si chiede che uso venga fatto di ciò che viene scaricato dalla rete, poichè il download non corrisponde necessariamente all'utilizzo e alla lettura.

Altro indicatore bibliometrico basato sull'analisi dell'uso dei link è il *web impact factor* (Ingwersen 1998) che si ricava dal numero di pagine di un sito web che ricevono link da altri siti, diviso per il numero di pagine web pubblicate nel sito stesso. Anche in questo caso i vantaggi sono legati alla flessibilità dello strumento e al fatto che si tratta di una metrica già applicabile, mentre per quanto riguarda i limiti essi sono relativi principalmente al livello di approssimazione delle misure fornite e al fatto che non può essere considerato un indicatore di qualità (Noruzi 2006).

2.5. Conclusioni

In Italia le comunità scientifiche afferenti alle aree delle Scienze umane e sociali, corrispondenti alle aree CUN 10-14, hanno cominciato solo recentemente ad occuparsi di metodologia della valutazione della ricerca, ritenendo che la *peer review* sia l'unico strumento di valutazione possibile.

Eppure gli indicatori bibliometrici sono diventati molto popolari, malgrado i loro limiti, perchè offrono soluzioni semplici a problemi complessi, essendo meno costosi e di più facile implementazione della *peer review*. Gli indici bibliometrici permettono però solo di cogliere alcuni aspetti dell'impatto di una pubblicazione risultando deboli rispetto alla valutazione della qualità intrinseca.

Dal punto di vista degli strumenti la soluzione va ricercata in un approccio misto, che valuti la produttività di uno studioso in base ad indicatori quantitativi (indicatore di produttività del singolo ricercatore in un range temporale dato, indicatore di produttività del singolo rispetto al settore disciplinare di appartenenza, indicatore di produttività del singolo rispetto alla struttura di appartenenza, e così via) e che si affidi invece alla *peer review* per un giudizio di ordine qualitativo.

Ma questo non basta: per le scienze umane non è sufficiente applicare i parametri da sempre utilizzati nelle scienze dure, ad esempio estendendo ad esse la copertura dei principali database ma è opportuno creare strumenti *ad hoc* per esaltarne le specificità. In questo senso, come detto, le biblioteche possono giocare un ruolo di primaria importanza.

Dato questo obiettivo la strada da percorrere sembra ormai chiara e di seguito se ne sintetizzano i passaggi:

- in primo luogo è necessario favorire una raccolta di dati di qualità – anche con l'aiuto dei bibliotecari – promuovendo la diffusione di archivi istituzionali negli atenei che siano interfacciati con l'anagrafe certificata della ricerca (ANPrePS), che avrebbe dovuto essere la base dell'attuale esercizio di valutazione VQR e che ANVUR ha individuato come una delle priorità per i prossimi anni¹³;

¹³ ANPrePS (Anagrafe nominativa dei professori ordinari e associati e dei ricercatori) prevista da una norma del 2009 (art. 3 bis legge 1/2009) contenente per ciascun soggetto l'elenco delle pubblicazioni scientifiche prodotte.

- è opportuno capire come funzionano realmente i diversi s.s.d. afferenti alle varie aree. Va ricordato che all'interno della stessa area, pur in presenza di una relativa omogeneità di metodo, i singoli settori si caratterizzano per esiti diversi e per diverse pratiche citazionali. Ha senso adottare per un insieme eterogeneo di discipline il medesimo criterio di misura sulla base dell'appartenenza al complesso delle scienze umane? Siamo certi che una scelta del genere non finisca per danneggiare una parte rilevante delle discipline umanistiche?
- infine è importante cominciare a sperimentare l'applicazione di sistemi misti di misurazione e valutazione dei contributi depositati grazie alla presenza dei full-text. Un buon sistema di valutazione dovrebbe sempre prevedere una triangolazione fra diverse tecniche di analisi e dovrebbe essere applicato con la consapevolezza che, laddove si usino parametri quantitativi, essi divengono attendibili quando oggetto della valutazione è un'aggregazione di soggetti (dipartimenti etc.) e non i singoli individui.

2.6. Bibliografia

- ARCHAMBAULT E. ET AL. (2006). *Benchmarking Scientific Output in The Social Sciences and Humanities: the Limits of Existing Databases*. «Scientometrics». vol. 68 n. 3, pp. 329–342.
- BACCINI A. (2010). *Valutare la ricerca scientifica. Uso e abuso degli indicatori bibliometrici*, il Mulino.
- BANFI A. (2012). *Apples and oranges? Spunti per una discussione sulla valutazione della ricerca nelle scienze umane e sociali*. «Roars. Return On Academic Research». URL= <http://www.roars.it/online/wpcontent/uploads/2012/08/Apples-and-oranges.pdf>. [ultima visita 15.1.2013].
- CASSELLA M. (2010). *Social peer-review e scienze umane, ovvero "della qualità nella Repubblica della scienza"*. «JLIS», vol.1, n.1. URL= <http://leo.cilea.it/index.php/jlis/article/download/30/35>. [ultima visita 15.1.2013].
- CASSELLA M. BOZZARELLI O. (2011). *Nuovi scenari per la valutazione della ricerca tra indicatori bibliometrici citazionali e metriche alternative nel contesto digitale*. «Biblioteche oggi», vol. 29, n. 2, p. 66-78.
- COMBA V. (2003). *La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives*. «Bollettino AIB», vol. 43, n. 1, pp. 65-76.
- DE BELLIS N. (2009). *Bibliometrics and citation analysis: from the Science citation index to cybermetrics*, Scarecrow Press.
- FAGGIOLANI C., SOLIMINE G. (2012). *La valutazione della ricerca, la bibliometria e l'albero di Bertoldo*. «AIB Studi», vol. 52, n.1. pp. 57-63.
- FIGÀ TALAMANCA A. *L'impact factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*. IV Seminario Sistema Informativo Nazionale per la Matematica SINM 2000: un modello di sistema informativo nazionale per aree disciplinari, Lecce, 2 ottobre 2000. URL= <http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>. [ultima visita 15.1.2013].
- GALIMBERTI P. (2012). *Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia*. «JLIS.it». vol.3, n.1. URL= <http://leo.cilea.it/index.php/jlis/article/download/5617/5771>. [ultima visita 15.1.2013].

- EAD. (2010). *Verso un nuovo scenario per la valutazione della ricerca: potenzialità dell'Open Access e limiti imposti dal contesto*. «JLIS.it», vol. 1, n. 1, pp. 87-110. URL= <http://leo.cilea.it/index.php/jlis/article/view/16/23>. [ultima visita 15.1.2013].
- GILBERT N. G. (1977). *Referencing as Persuasion*. «Social Studies of Science», vol. 7 n. 1, pp. 113-122.
- GUERRINI M. (2009) *Nuovi strumenti per la valutazione della ricerca scientifica: il movimento dell'open access e gli archivi istituzionali*, «Biblioteche oggi», vol. 27, n. 8, pp. 7-17.
- HELLQVIST B. (2010). *Referencing in the Humanities and Its Implications for Citation Analysis*. «Journal of the American Society for Information Science and Technology». vol. 61 n. 2, pp. 310-318.
- HICKS D. (2004). *The Four Literatures for Social Science*. In H. Moed, W. Glanzel, U. Schmoch, a c.di, *Handbook of Quantitative Science and Technology Research*. Dordrecht. Kluwer Academic Publisher, 2004, pp. 473-496.
- HICKS, D., WANG, J. (2010). *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*. A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF, March 2010. URL= http://globalhighered.files.wordpress.com/2010/07/esf_report_final_100309.pdf. [ultima visita 15.1.2013].
- INGWERSEN P. (1998). *The calculation of web impact factors*, «Journal of Documentation», vol. 54, n. 2, pp. 236-243.
- MERTON R. (1988). *The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property*. «Isis», vol. 79 pp. 606-623, URL= <http://www.garfield.library.upenn.edu/merton/matthewii.pdf> [ultima visita 15.1.2013].
- MOED H. (2008). *Research Assessments in Social Sciences and Humanities*, URL= <http://www2.lingue.unibo.it/evaluationinthehumanities/Research%20Assessment%20in%20Social%20Sciences%20and%20Humanities.pdf>. [ultima visita 15.1.2013].
- NEDERHOF A.J. (2006). *Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: a Review*. «Scientometrics», vol. 66 n. 1, pp. 81-100.
- NORUZI A. (2006). *The web impact factor: a critical review*. «The Electronic Library», vol. 24, n. 4, pp. 490-500.

- SERRAI A. (1984), *Dai "loci communes" alla bibliometria*, Bulzoni.
- TAMMARO A. (2001). *Qualità della comunicazione scientifica. Gli inganni dell' impact factor e l'alternativa della biblioteca digitale*, «Biblioteche oggi», vol. 19, n. 7, pp. 104-107.
- TARANTINO E. (2006). *Troppo o troppo poco? Web of science, Scopus, Google scholar: tre database a confronto*. «Bollettino AIB», vol.46, n.1/2, pp. 23-34, URL= <http://bollettino.aib.it/article/view/5132/489>. [ultima visita 15.1.2013].
- TORRES SALINAS D., MOED H. (2009). *Library Catalog Analysis as a Tool in Studies of Social Sciences and Humanities: an Exploratory Study of Published Book Titles in Economics*, «Journal of Informetrics», vol. 3, n. 1, pp. 9-26, URL= eprints.rclis.org/15705. [ultima visita 15.1.2013].

3. Biblioteche digitali tra modellazione, gestione e valutazione

Maritella Agosti, Nicola Ferro, Gianmaria Silvello¹

Abstract. Le biblioteche digitali e i sistemi di gestione di biblioteche digitali operano in contesti eterogenei e in rapida evoluzione. Ne consegue che i sistemi che vengono ideati ed utilizzati devono essere progettati per essere dinamici e in grado di gestire l'interoperabilità con altri sistemi per favorire la fruizione dei contenuti digitali da parte di diverse categorie di utenti. Per raggiungere questi obiettivi di dinamicità e interoperabilità i sistemi di biblioteche digitali devono far riferimento a modelli di qualità per gestire i contenuti in modo consistente. Per questo si illustra un modello di qualità che può essere adottabile per la conservazione della qualità di una biblioteca digitale nel tempo. Da ultimo si presentano gli aspetti fondamentali della valutazione sperimentale, perché, utilizzando i metodi propri della valutazione sperimentale, si attua un circolo virtuoso che tiene conto delle varie caratteristiche utili ad attuare sistemi orientati alla soddisfazione degli utenti finali.

Parole chiave: biblioteca digitale, sistema di gestione di biblioteca digitale, modello di rappresentazione e gestione dell'informazione, qualità, modello di qualità, valutazione sperimentale, dato sperimentale.

¹ Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova, Padova, Italia. e-mail: {agosti,ferro,silvello}@dei.unipd.it.

3.1. Introduzione

Con il termine 'biblioteca digitale' si fa riferimento ai metodi e agli strumenti dell'informatica utili e necessari alla gestione di materiale tradizionale e digitale custodito nelle biblioteche, negli archivi e in altre istituzioni di conservazione e fruizione per il pubblico, come ad esempio i musei. Infatti le biblioteche digitali e i corrispondenti sistemi di gestione di contenuti mirano ad essere sia i depositi di varie forme di conoscenza sia i mezzi attraverso cui i cittadini possono accedere, discutere, valutare ed arricchire diversi tipi di contenuti informativi, anche con riferimento alla conservazione, alla valorizzazione e alla diffusione del patrimonio culturale (Agosti 2012). Infatti i sistemi di gestione di biblioteche digitali sono in grado di preservare e gestire una o più biblioteche digitali permettendone la fruizione ai possibili utenti interessati.

Le biblioteche digitali e i sistemi in grado di mantenerle e renderle disponibili rappresentano un punto d'incontro di molti ambiti disciplinari come la gestione dei dati e delle basi di dati, del reperimento dell'informazione, del Web, dell'archivistica, della bibliografia, della biblioteconomia, dei sistemi informativi, dell'interazione uomo-macchina e della conservazione degli oggetti digitali (Agosti Ferro 2010).

Chi opera per la realizzazione o per l'utilizzazione di sistemi di gestione di biblioteche digitali si trova ad interagire con un contesto eterogeneo e in rapida evoluzione dove i sistemi vengono ideati per essere dinamici e in grado di gestire l'interoperabilità con altri sistemi in modo tale da favorire la fruizione di contenuti digitali da parte di diverse categorie di utenti. Per raggiungere questi obiettivi di dinamicità e interoperabilità i sistemi di biblioteche digitali devono essere ideati e realizzati facendo riferimento a dei modelli in grado di permettere simili realizzazioni e che fanno riferimento a un modello di qualità che evita che la gestione dei contenuti possa risultare problematica nel tempo. Per questo nel seguito si illustrano gli elementi essenziali di modellazione e gestione di biblioteche digitali ai quali è possibile fare riferimento. Poi si illustra un modello di qualità che può essere adottato per conservare la qualità della biblioteca digitale nel tempo. Da ultimo si presentano gli aspetti fondamentali della valutazione sperimentale nel contesto delle biblioteche digitali per illustrare come questi aspetti siano utili per la ideazione dei modelli di sistemi con le caratteristiche alle quali si è già fatto riferimento, quindi modelli che

permettono di realizzare sistemi dinamici e interoperabili. In pratica, si illustra come, affrontando la ideazione e realizzazione di modelli di sistemi di biblioteche digitali con le caratteristiche già citate che fanno riferimento ad un modello di qualità e utilizzando i metodi propri della valutazione sperimentale, si possa attuare un circolo virtuoso che tiene conto delle varie caratteristiche utili ad attuare sistemi orientati alla soddisfazione di diverse categorie di utenti finali.

3.2. Modellare e gestire una biblioteca digitale

Le problematiche connesse all'ideazione e realizzazione di sistemi dinamici e interoperabili riguardano in particolare la necessità di creare nuovi modelli per l'automazione dei processi di rappresentazione ed elaborazione dell'informazione riguardante le specifiche risorse del patrimonio culturale che si desidera rappresentare e gestire in forma digitale. A seconda della tipologia delle risorse culturali, di volta in volta di specifico interesse, il processo d'ideazione di un nuovo modello di gestione dell'informazione deve avvenire da una collaborazione effettiva che si deve instaurare fra gli esperti dello specifico dominio del patrimonio culturale e gli esperti informatici di sistemi di elaborazione delle informazioni. Infatti gli esperti dello specifico dominio del patrimonio culturale – come ad esempio esperti di storia dell'arte, di archivistica, di biblioteconomia, di archeologia, di linguistica, di storia – conoscono la storia e i metodi del loro specifico dominio e gli esperti di sistemi di elaborazione dell'informazione conoscono la storia e i metodi propri dell'informatica. Solamente grazie ad un rapporto sinergico di elaborazione di nuove soluzioni metodologiche gli esperti dei due settori possono prendere in considerazione i requisiti funzionali e di gestione dell'informazione degli utenti del dominio e ideare insieme nuovi metodi e soluzioni.

Questo processo d'ideazione non si limita, quindi, all'analisi dei requisiti d'utente condotta in modo isolato da parte degli informatici esperti di sistemi di gestione dell'informazione, ma da parte degli esperti informatici insieme agli specialisti dello specifico settore del patrimonio culturale d'interesse per identificare gli aspetti peculiari del settore e le caratteristiche primitive di gestione dell'informazione che occorre tenere presente per la ideazione e formulazione di un nuovo modello di gestione (Agosti 2008).

Una volta che questo nuovo modello è stato ideato e formalizzato è possibile progettare un corrispondente nuovo sistema di gestione dell'informazione; a questo proposito l'informatica diventa allora una delle culture necessarie perché grazie a un approccio innovativo di gestione, permette di produrre nuove conoscenze che prima non potevano essere rappresentate ed elaborate. L'ideazione di nuovi modelli e sistemi risulta necessario perché gli aspetti della realtà che si prendono in considerazione per una gestione automatica dell'informazione sono più complessi di quelli che venivano affrontati in precedenza, quindi devono essere ideati nuovi metodi e sistemi di gestione dell'informazione a mano a mano che la complessità degli aspetti della realtà da affrontare e gestire aumenta; questo processo può avvenire mediante l'analisi e l'ideazione condotta insieme agli esperti del dominio culturale specifico in un processo iterativo simile a quello rappresentato in Figura 3.1

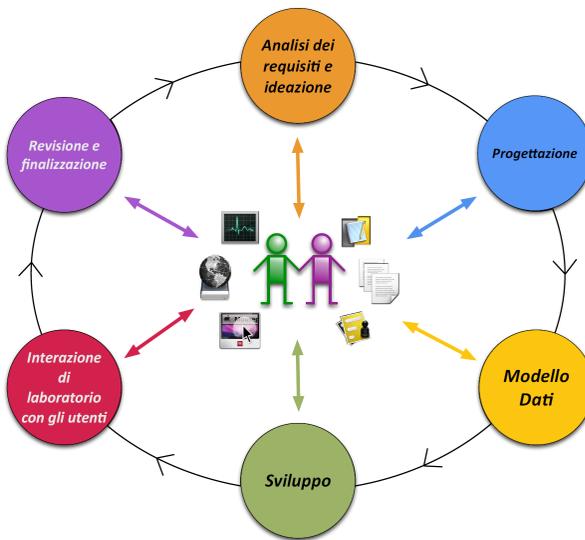


Fig.3.1. Approccio progettuale.

Per ideare i metodi e i componenti che affrontano e realizzano le azioni presentate in Figura 3.1, il primo passo da compiere è mettere in evidenza i diversi livelli di possibile intervento, suddividendo il problema complessivo in sottoproblemi, come sempre accade quando si deve ideare una soluzione scientifica per risolvere un problema

complesso. Nel contesto dei sistemi di elaborazione dell'informazione tali sottoproblemi vengono identificati e messi in evidenza prendendo in considerazione l'interazione che il sistema avrà con l'utente finale (Agosti 2012b). Infatti, il sistema viene utilizzato dall'utente finale attraverso un'interfaccia grafica che presenta le caratteristiche e le funzioni utilizzabili, ma si basa su sottosistemi interni di elaborazione delle caratteristiche e delle funzioni di gestione di dati che rappresentano l'informazione fruibile e gestibile dall'utente. Un sistema di questo tipo viene rappresentato solitamente mediante un'articolazione a tre livelli – come schematizzato in Figura 3.2 – che corrispondono:

- al livello esterno, o livello di presentazione dei servizi, dove avviene l'interazione con l'utente finale grazie ai servizi di elaborazione dell'informazione forniti dal modello ideato e realizzati dal sistema fornito;
- al livello intermedio, o livello di logica dei servizi, dove avvengono le elaborazioni dell'informazione di specifico interesse dell'applicazione che si rende disponibile all'utente;
- al livello interno, o livello di logica dei dati, dove l'informazione di interesse viene rappresentata per essere mantenuta nel tempo a disposizione dell'utente anche con l'ausilio di dispositivi di memorizzazione permanente.

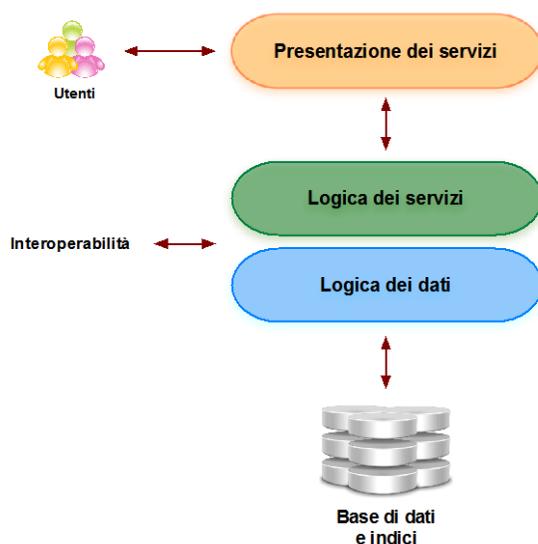


Fig. 3.1. I tre livelli di rappresentazione e gestione dell'informazione.

Le funzionalità fornite grazie alla risoluzione dei problemi affrontati, vengono rese disponibili agli utenti mediante un'interfaccia grafica che in genere scherma gli effettivi metodi elaborati internamente dal sistema a livello di logica dei servizi. Il livello interno realizza i metodi utili alla gestione e al mantenimento nel tempo dei dati d'interesse servendosi degli strumenti di ausilio alla memorizzazione dei dati, quali sono gli indici che contribuiscono ad una gestione efficiente dei dati.

Per affrontare alcuni dei sottoproblemi, in cui il problema considerato viene suddiviso, può capitare di poter disporre di metodi e di soluzioni già resi disponibili in precedenza. Questo capita più di frequente per il livello interno dove l'informazione d'interesse viene rappresentata sotto forma di dati; infatti, un modello di rappresentazione dell'informazione può essere stato ideato e attuato con una soluzione metodologica così valida e coerente, da mantenere la sua validità molto a lungo. Un esempio di un modello di rappresentazione e gestione di dati a livello interno che costituisce ancora oggi una soluzione metodologica estremamente solida è il modello di dati relazionale ideato alla fine degli anni 1960/1970 da Edgar Frank (normalmente chiamato Ted) Codd (Codd 1970) e poi successivamente elaborato e completato dallo stesso Codd (Codd 1979, Codd 1982) e dalla comunità scientifica internazionale.

Visto che il livello di logica dei dati si basa in gran parte su un modello di dati che in genere è di tipo relazionale, i livelli dove si collocano la maggior parte dei nuovi modelli da ideare per affrontare la rappresentazione e la gestione di nuovi aspetti della realtà, sono quelli relativi alla logica dei servizi e alla loro presentazione. Un esempio significativo di un progetto, per il quale si è ideato prima un modello e poi si è sviluppato un sistema con una metodologia coerente a quanto sopra illustrato, è il 'Sistema Informativo Archivistico Regionale' (SIAR): un progetto avviato e poi sostenuto dalla Regione del Veneto per lo sviluppo di un sistema di gestione e condivisione distribuita di informazioni di natura archivistica.²

In questo contesto, si è definito un modello dei dati basato sulla teoria degli insiemi chiamato NESTOR (*NEsted SeTs for Object*

² <http://www.regione.veneto.it/Servizi+alla+Persona/Cultura/Beni+culturali/Archivi/Sistema+Informativo+Archivistico+Regionale.htm>.

hieRarchies), descritto in (Ferro Silvello 2011); conseguentemente si è sviluppato un sistema innovativo di rappresentazione e gestione dell'informazione, illustrato in (Agosti *et al.* 2011), che ora è in fase avanzata di test in vista del rilascio al pubblico. Sia il modello dei dati, che il sistema sviluppato, sono coerenti con quanto presentato sopra; infatti, il sistema è il risultato di una effettiva sinergia tra informatici e archivisti attivata per mettere a frutto competenze diverse ma complementari per affrontare in modo proprio il problema considerato. Inoltre, questo processo ha messo e mette l'utente al centro del processo di ideazione, progettazione e sviluppo del sistema di gestione dell'informazione.

Per concludere questa parte del contributo cerchiamo di mettere in luce il modo in cui gli informatici possono contribuire all'ideazione di un nuovo sistema informativo da automatizzare. Va sottolineato, allora, che l'interscambio fra gli esperti informatici e gli specialisti dello specifico settore del patrimonio culturale d'interesse non deve essere limitato ad un mero livello tecnologico, ma deve avvenire a livello scientifico e di saperi e richiede un grande lavoro di collaborazione. L'informatica è una delle culture che deve essere parte di questo nuovo fare sistema, per creare un approccio innovativo che abbia come obiettivo la produzione di nuovi sistemi in grado di incorporare e gestire anche nuova conoscenza.

3.3. Come definire e valutare la 'qualità' delle biblioteche digitali

Il modello per definire e valutare la qualità delle biblioteche digitali e dei sistemi di gestione di biblioteche digitali, sviluppato nell'ambito del 'Modello di riferimento DELOS' (Candela *et al.* 2007), è quello che ad oggi può essere considerato il più avanzato e da utilizzare perché comprende tutte le caratteristiche già indicate in (Agosti *et al.* 2007) a questo scopo e si basa su schemi e modelli solidi che erano già stati sviluppati e proposti, ad esempio in (Fuhr *et al.* 2001).

Il modello di qualità delle biblioteche digitali e dei sistemi di gestione di biblioteche digitali, che è parte del Modello di riferimento DELOS, deve la sua validità proprio alla completezza di questo modello, che fornisce la possibilità sia di definire la qualità in un modo consistente per le diverse entità che compongono una biblioteca

digitale, che la possibilità di effettuare la modellizzazione ad un livello appropriato di astrazione. Questo modello permette anche di non affrontare la modellazione di dettagli non necessari ma, allo stesso tempo, assicura che il modello possa essere applicato ad un'ampia gamma di casi concreti, queste caratteristiche fanno sì che questo modello di qualità costituisca un avanzamento rispetto agli schemi e ai modelli definiti in precedenza e, ad oggi, non risulta essere ancora stato superato.

La caratteristica fondamentale di questo modello è che non si concentra solo sulla definizione dei parametri e degli indicatori da esaminare per valutare la qualità di una biblioteca digitale, ma mette in relazione la nozione di qualità e le sue molteplici caratteristiche con le altre entità e concetti che costituiscono una biblioteca digitale e così permette di determinare quali sono gli effetti di reciproca influenza. Per uno studio approfondito del modello di qualità e delle sue caratteristiche formali si rimanda il lettore a (Ferro 2010).

3.4. La valutazione sperimentale e le biblioteche digitali

Nel caso del patrimonio culturale, gli utenti desiderano sistemi di biblioteche digitali che permettano di interrogare simultaneamente cataloghi di biblioteche, archivi e musei per accedere a libri, scansioni di manoscritti, riproduzioni di quadri, opere musicali, documenti audiovisivi, e così via. Questo è il caso, ad esempio, di Europeana³, un progetto che mira alla creazione della biblioteca digitale europea in cui confluiscono i materiali provenienti da biblioteche, musei e archivi (anche audiovisivi) al fine di valorizzare e rendere accessibile il patrimonio culturale europeo attraverso un portale unico. Per conseguire questo ambizioso obiettivo diviene necessario ideare e sviluppare modelli di dati avanzati, capaci di mediare tra le diverse modalità di descrizione del patrimonio culturale che caratterizzano le varie istituzioni che devono mantenerlo, così come progettare e realizzare sistemi in grado di implementare tale modello di dati e offrire funzionalità di ricerca basate su esso in modo efficiente e scalabile.

³ <http://www.europeana.eu/>.

Di conseguenza, la progettazione e sviluppo dei sistemi di accesso all'informazione, che, nel caso dei sistemi di biblioteche digitali, costituiscono il componente che permette l'accesso e il recupero delle informazioni gestite, diviene un'attività sempre più complessa e che richiede una costante verifica per assicurare che essi siano conformi ai requisiti specificati e che garantiscano le prestazioni attese dagli utenti. In questo processo di progettazione e sviluppo, la valutazione sperimentale di tali sistemi svolge un ruolo centrale. Per 'valutazione sperimentale' si intende l'attività volta a verificare e quantificare le effettive prestazioni di tali sistemi rispetto a dei 'campioni' di riferimento.

La valutazione sperimentale è un'attività di supporto e guida allo sviluppo dei sistemi di accesso all'informazione. Essa si ispira ai metodi e alle pratiche del procedimento scientifico in quanto, grazie all'utilizzo di collezioni di dati e scenari di valutazione condivisi, rende possibile confrontare i sistemi di accesso all'informazione, comprendere e analizzarne le prestazioni e perseguire e dimostrare i progressi ottenuti. La valutazione sperimentale viene usualmente condotta nel contesto di campagne internazionali di valutazione su larga scala al fine di favorire la condivisione delle risorse, la comparabilità e il confronto tra i risultati sperimentali, le tecniche e gli algoritmi, e la formazione di una comunità con le competenze multi-disciplinari necessarie per sviluppare i sistemi di accesso all'informazione anche multilingua e multimediale. Infatti, i componenti di accesso alle informazioni utilizzati all'interno dei sistemi di biblioteche digitali sempre più spesso necessitano di capacità di gestione di contenuti in lingue diverse e provenienti da supporti originali diversi, quindi di tipo multimediale.

Le maggiori campagne di valutazione sperimentale a livello internazionale sono la *Text REtrieval Conference* (TREC)⁴ organizzata dal *National Institute for Standards and Technology* (NIST) negli Stati Uniti, la *The CLEF Initiative* (dove CLEF sta per *Conference and Labs of the Evaluation Forum*)⁵ in Europa, e il *NII-NACSIS Test Collection for IR systems* (NTCIR)⁶ organizzato dal *National Institute of Informatics* (NII) in Giappone e Asia. Inoltre, stanno sorgendo nuove iniziative di

⁴ TREC, <http://trec.nist.gov/>.

⁵ CLEF, <http://www.clef-initiative.eu/>.

⁶ NTCIR, <http://research.nii.ac.jp/ntcir/>.

questo genere per dare voce e supporto a nuove comunità e per indirizzare esigenze specifiche, come ad esempio il *Forum for Information Retrieval Evaluation (FIRE)*⁷ in India. Tutte le campagne di valutazione seguono un paradigma comune, chiamato 'paradigma di Cranfield', sviluppato nella seconda metà degli anni sessanta del secolo scorso e che fa ricorso all'uso di collezioni sperimentali per quantificare l'efficacia di un sistema di accesso all'informazione (Harman 2011).

Una collezione sperimentale consiste:

- di una collezione di documenti rappresentativi, per argomento e numerosità, del dominio in cui il sistema di accesso all'informazione dovrà operare,
- di una collezione di esigenze informative, da cui si derivano le interrogazioni da porre al sistema, e
- di una serie di giudizi di rilevanza che, per ogni interrogazione, indicano quali documenti della collezione sono rilevanti e quali non lo sono.

Utilizzando una collezione sperimentale, è possibile determinare quali dovrebbero essere i risultati 'ideali' forniti da un sistema di accesso all'informazione in risposta ad una data interrogazione e si ha quindi la possibilità di confrontare con essi i risultati effettivi prodotti da un dato sistema, quando gestisce quella specifica collezione sperimentale, determinandone così le prestazioni.

In questo contesto, le campagne di valutazione sperimentale forniscono un contributo significativo all'avanzamento dello stato dell'arte della ricerca nel settore e all'innovazione industriale.

3.4.1. La gestione e conservazione dei dati prodotti dalla valutazione

Durante la loro vita, le campagne di valutazione sperimentale su larga scala hanno prodotto e producono una grande quantità di dati scientifici di estremo valore. Non solo essi rappresentano il prodotto di sistemi che hanno richiesto molti anni di lavoro per la loro ideazione, progettazione e sviluppo, ma questi dati sono anche spesso difficili da riprodurre, sia per la complessità dei sistemi stessi, sia per

⁷ FIRE, <http://www.isical.ac.in/~fire/>.

il costo delle infrastrutture hardware su cui essi funzionano, che non sono alla portata di tutte le organizzazioni o dei gruppi di ricerca. Questi dati sperimentali forniscono le basi per tutta la produzione scientifica a venire e per lo sviluppo dei sistemi futuri; inoltre, è estremamente importante che questi dati siano facilmente accessibili, verificabili e utilizzabili per nuovi confronti qualora fosse necessario.

Tuttavia, minore attenzione è stata posta fino ad oggi alla modellazione, gestione, cura, accesso, preservazione e riuso di tali dati scientifici, anche se la loro generale importanza è stata evidenziata da diverse importanti istituzioni come, ad esempio, la Commissione Europea (CEC 2007), il *National Scientific Board* (NSB) (NSB 2005) negli Stati Uniti, e l'*Australian Working Group on Data for Science* (WGDS 2006) in Australia.

Occorre quindi ideare, sviluppare e fornire un'infrastruttura e un ambiente unificati in cui collezionare i dati scientifici e gli strumenti a supporto della valutazione sperimentale per diverse comunità di utenti, singoli ricercatori, e organizzazioni commerciali che si occupano di progettare, realizzare, migliorare, adottare e utilizzare i sistemi di accesso all'informazione.

Il primo passo in questa direzione è rappresentato dalla ideazione e realizzazione del sistema di biblioteca digitale per la gestione dei dati scientifici, prodotti dalla valutazione sperimentale, chiamato DIRECT⁸ – *Distributed Information Retrieval Evaluation Campaign Tool* – che, come illustrato in (Agosti Ferro 2009) e in (Ferro *et al.* 2011), ha le seguenti caratteristiche salienti:

- la modellazione delle entità coinvolte nello spazio informativo della valutazione sperimentale, dei loro attributi e delle relazioni tra loro avviene, per la prima volta, mediante uno schema concettuale;
- il sistema fornisce le basi e gli strumenti per sviluppare formati di metadati che consentono di descrivere il significato dei dati scientifici che vengono gestiti, di facilitare la loro condivisione, scambio e riuso nel tempo;
- adotta un meccanismo di identificazione univoca che consente la citazione esplicita dei dati scientifici, ad esempio negli articoli dove vengono utilizzati, e ne favorisce un accesso facilitato;
- gestisce i diversi aspetti di una campagna di valutazione sperimentale

⁸ <http://direct.dei.unipd.it/>.

come, ad esempio, la gestione delle collezioni di documenti, la sottomissione degli esperimenti, il calcolo delle misure di prestazione. La necessità, da parte del settore scientifico di riferimento, di avere a disposizione un sistema di biblioteca digitale quale DIRECT, è dimostrata dall'insieme di dati sperimentali che sono già gestiti attraverso questa infrastruttura. Infatti, ad oggi (Ferro 2012), DIRECT gestisce più di 8,6 milioni di documenti in 16 lingue diverse e scritti con 4 differenti alfabeti (latino, cirillico, arabo e cinese); mantiene 4.200 esperimenti che ammontano a più di 300 milioni di elementi e sono stati prodotti da 430 partecipanti in 41 nazioni diverse; contiene più di 3,2 milioni di giudizi di rilevanza per più di 10.000 interrogazioni prodotti da oltre 270 valutatori in 19 nazioni diverse; dà accesso ad oltre 27 milioni di misurazioni e statistiche descrittive sugli esperimenti gestiti e a più di 50.000 grafici ed analisi statistiche.

DIRECT si fonda su un'architettura flessibile e modulare (cfr. Fig. 3.3), basata su servizi Web, che consente un accesso e una manipolazione dei dati gestiti indipendente dalla specifica applicazione che si desidera realizzare. In questo modo, facendo leva su questo sistema, si possono realizzare diverse applicazioni adatte a diverse esigenze e categorie di utenza.

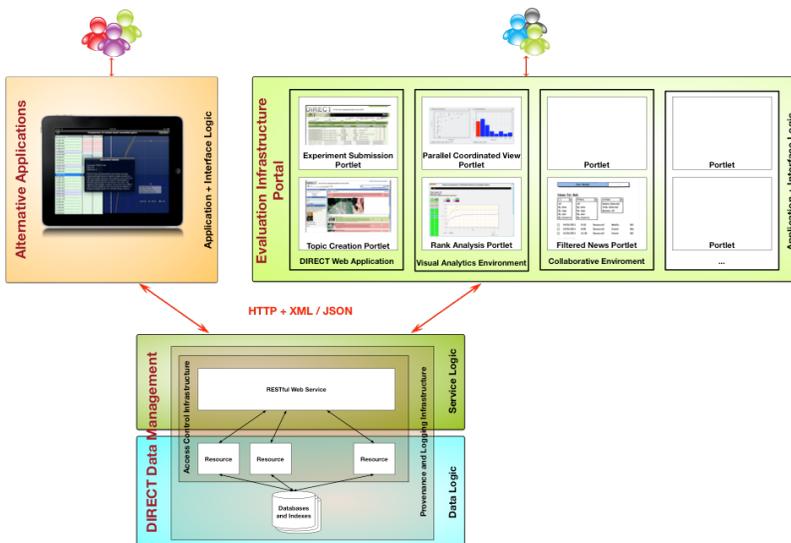


Fig. 3.2. Architettura del sistema DIRECT.

3.4.2. Funzionalità avanzate per l'accesso ai dati sperimentali

Avere a disposizione un'infrastruttura per la gestione e l'accesso ai dati sperimentali, offre l'opportunità di sviluppare applicazioni avanzate su di essa, che favoriscono la collaborazione, mantenimento e interazione con i dati sperimentali. Una possibilità interessante è quella di consentire l'inserimento, da parte degli utenti, di annotazioni (Agosti Ferro 2008) sui dati gestiti (cfr. Fig. 3.4).

Le annotazioni possono consentire agli utenti sia di collaborare, utilizzando i dati sperimentali (ad esempio fornendo spiegazioni e discutendo le prestazioni di un dato sistema), sia di migliorare la qualità dei dati mantenuti, segnalando possibili errori o inconsistenze. Tutto ciò favorisce il mantenimento dei dati nel tempo e la possibilità di poterli interpretare e comprendere correttamente anche a distanza di anni.

L'infrastruttura consente anche di esplorare l'utilizzo di dispositivi alternativi per navigare, manipolare ed interagire con i dati sperimentali. In Fig. 3.5 viene illustrato un primo prototipo di applicazione per *tablet* (in questo caso l'applicazione ottimizzata per Apple iPad⁹) finalizzata a confrontare i risultati sperimentali con quelli ottimali ed ideali. Un possibile scenario di utilizzo di una tale applicazione è quello di un gruppo di lavoro che stia sviluppando un sistema di accesso alle informazioni e necessiti di discutere ed analizzarne le prestazioni per capire meglio come sta procedendo lo sviluppo. Avere a disposizione un'applicazione su dispositivo *tablet* o 'mobile' che permette di manipolare, esplorare, ed interagire con i dati sperimentali con una modalità di interazione più semplice, può rappresentare un modo efficace per facilitare la comunicazione all'interno del gruppo di lavoro oppure condurre una riunione e far così progredire il lavoro.

⁹ <http://www.apple.com/it/ipad/>.

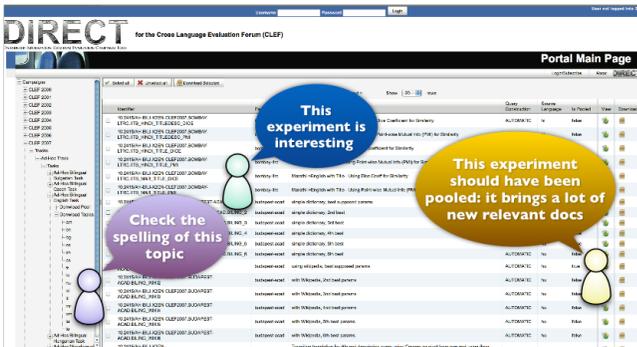


Fig. 3.3. Annotazioni relative a dati sperimentali.

Un ulteriore scenario di utilizzo di questa applicazione è lo studio di un articolo scientifico che descriva un sistema e le sue prestazioni e la possibilità di 'cliccare' direttamente su un grafico ed iniziare ad esplorare interattivamente i dati descritti nell'articolo (cfr. Fig. 3.6). In Figura 3.6 si sta utilizzando l'applicazione iBooks¹⁰, già pre-installata su ogni tablet Apple di tipo iPad, e l'utente che sta interagendo con l'applicazione seleziona il link 'topic 351' mostrato in azzurro. A questo punto si apre automaticamente l'applicazione mostrata in Figura 3.5 e l'utente può iniziare ad interagire con i dati presentati e discussi nell'articolo che sta leggendo.

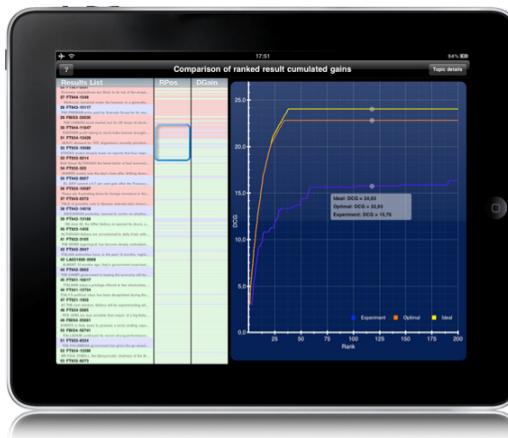


Fig. 3.4. Esempio di applicazione per *tablet* per l'esplorazione dei risultati sperimentali.

¹⁰ <http://www.apple.com/it/apps/ibooks/>.



Fig. 3.5. Esempio di lettura di un articolo scientifico e consultazione dei dati sperimentali corrispondenti.

3.5. Ringraziamenti

Il lavoro riportato in questo articolo è stato parzialmente sostenuto dal progetto CULTURA (contratto n. 269973)¹¹ e dalla rete di eccellenza PROMISE (contratto n. 258191)¹² ambedue del settimo programma quadro della Commissione Europea.

The work reported has been partially supported by the CULTURA project (contratto n. 269973) and by the PROMISE network of excellence (contract n. 258191) project, as part of the 7th Framework Programme of the European Commission.

¹¹ <http://www.cultura-strep.eu/home>.

¹² <http://www.promise-noe.eu/home>.

3.6. Bibliografia

- AGOSTI M. (2008). *Information Access using the Guide of User Requirements*. In M. Agosti, a c. di, *Access through Search Engines and Digital Libraries*. Springer-Verlag, pp. 1-12.
- AGOSTI M. (2012). *Digital Libraries*. «Mondo Digitale», vol. 43, settembre 2012, pp. 1-13.
- AGOSTI, M. (2012b). *Una nuova generazione di sistemi di gestione e conservazione di risorse digitali del patrimonio culturale grazie al coinvolgimento di diverse comunità di utenti*. «Atti e Memorie dell'Accademia Galileiana di Scienze Lettere ed Arti in Padova, già Dei Ricoverati e Patavina». Parte II: Memorie della Classe di Scienze Matematiche, Fisiche e Naturali, vol. CXXIV, pp. 11-22.
- AGOSTI M., FERRO, N. (2008). *A Formal Model of Annotations of Digital Content*. «ACM Transactions on Information Systems», vol. 26, no 1, pp. 3-57.
- AGOSTI M., FERRO N. (2009). *Towards an Evaluation Infrastructure for DL Performance Evaluation*. In G. Tsakonas, C. Papatheodorou, a c. di, *Evaluation of Digital Libraries: An insight into useful applications and methods*, Chandos Publishing, Oxford, UK, pp. 93-120.
- AGOSTI M., FERRO N. (2010). *Interoperabilità tra sistemi di biblioteche digitali*. «DigItalia», vol. 5, no 1, pp. 95-112.
- AGOSTI M. ET AL. (2007). *Towards a Reference Quality Model for Digital Libraries*. In D. Castelli, E.A. Fox, a c. di, *Pre-proceedings of the First International Workshop on Foundations of Digital Libraries, 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2007)*, pp. 37-42.
- AGOSTI M. ET AL. (2011). *SIAR: A User-Centric Digital Archive System*. In *Digital Libraries and Archives, 7th Italian Research Conference, IRCDL 2011 Revised Papers*. Communications in Computer and Information Science, vol. 249, Springer-Verlag, pp. 87-99.
- CANDELA L. ET AL. (2007). *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*. ISTI-CNR at Gruppo ALI, Pisa, Italy
URL= http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.

- CODD E.F. (1970). *A relational model for large shared data banks*. «Communications of the ACM», vol. 13, no 6, pp. 377-387.
- CODD E.F. (1979). *Extending the database relational model to capture more meaning*. «ACM Transactions on Database Systems», vol. 4, no 4, pp. 397-434.
- CODD E.F. (1982). *Relational database: a practical foundation for productivity*. «Communications of the ACM», vol. 25, no 2, pp. 109-117.
- COMMISSION OF THE EUROPEAN COMMUNITIES (CEC) (2007). *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age: access, dissemination and preservation*. COMM (2008) 56 Final, February 2007.
- FERRO N. (2010). *Quality and Interoperability: The Quest for the Optimal Balance*. In I. Iglezakis, S. Kapidakis, T. Synodinou, a c. di, *E-Publishing and Digital Libraries: Legal and Organizational Issues*, IGI Global, USA, pp. 48-68.
- FERRO, N. (2012). *La gestione e conservazione dei dati scientifici prodotti dalla valutazione sperimentale dei motori di ricerca*. «Atti e Memorie dell'Accademia Galileiana di Scienze Lettere ed Arti in Padova, già Dei Ricoverati e Patavina». Parte II: Memorie della Classe di Scienze Matematiche, Fisiche e Naturali, vol. CXXIV, pp. 41-52.
- FERRO N., SILVELLO G. (2011). *The NESTOR Model: Properties and Applications in the Context of Digital Archives*. In G. Mecca, S. Greco, a c. di, *Proceedings of the 19th Italian Symposium on Advanced Database Systems (SEBD 2011)*. Università della Basilicata, Italy, pp. 274-285.
- FERRO N. ET AL. (2011). *Harnessing the Scientific Data Produced by the Experimental Evaluation of Search Engines and Information Access Systems*. *Procedia Computer Science*, vol. 4, pp. 740-749.
- FUHR N. ET AL. (2001). *Digital Libraries: A Generic Classification Scheme*. In P. Constantopoulos, I.T. Sølvsberg, a c. di, *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, LNCS 2163, Springer, Heidelberg, Germany, pp. 187-199.
- HARMAN D. (2011). *Information Retrieval Evaluation*. Morgan & Claypool Publishers.

NATIONAL SCIENCE BOARD (NSB) (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (NSB-05-40). National Science Foundation (NSF). URL=<http://www.nsf.gov/pubs/2005/nsb0540/>.

WORKING GROUP ON DATA FOR SCIENCE (WGDS) (2006). *From Data to Wisdom: Pathways to Successful Data Management for Australian Science*. Report to Prime Minister's Science, Engineering and Innovation Council (PMSEIC), URL= <http://www.innovation.gov.au/Science/PMSEIC/Documents/FromDatatoWisdom.pdf>.

4. Livelli descrittivi, relazioni e contesti di produzione nella Sapienza Digital Library

Cecilia Carloni, Maria Guercio¹

Abstract. L'ambiente informativo della Sapienza Digital Library intende raccogliere e rendere accessibili materiali di diversa natura e provenienza preservando i contesti di produzione, la profondità informativa e la complessità semantica dei patrimoni documentari. Una continua ricerca e sperimentazione degli strumenti tecnici e concettuali mira a costruire un'architettura dei contenuti che espliciti il contesto di produzione e le relazioni strutturali e semantiche, prevedendo quindi una descrizione a più livelli dei complessi documentari. Solo nuove forme di intermediazione tra sistemi informativi complessi e utenti possono veicolare le tracce culturali e i contesti che le hanno prodotte. Il modello di rappresentazione dei contenuti informativi distingue il piano dell'oggetto fisico da quello della rappresentazione digitale e semantica.

Parole chiave. biblioteche digitali, archivi digitali, collezioni, integrazione, convergenza digitale, interoperabilità, standard descrittivi.

¹ DigiLab Sapienza, Roma, Italia. e.mail: [cecilia.carloni](mailto:cecilia.carloni@uniroma1.it), [maria.guercio](mailto:maria.guercio@uniroma1.it)@uniroma1.it.

4.1. Introduzione

Il progetto Sapienza Digital Library (d'ora in avanti SDL) nasce con la missione di aggregare e rendere accessibile la documentazione di diversa natura e provenienza prodotta sia dai gruppi di ricerca operanti all'interno dell'Università, sia da enti, famiglie e persone esterne alla comunità accademica. Si intende quindi sviluppare un'architettura informativa che possa accogliere materiali diversi per tipologia (libri, periodici, spartiti musicali, fotografie, materiale multimediale, film, filmati audio e video, registrazioni sonore, ricostruzioni 3D di luoghi, edifici e territori) e provenienza, in quanto riconducibili a diverse modalità di produzione e sedimentazione di fonti documentarie e del patrimonio culturale -archivi, biblioteche, musei- nonché a differenti domini metodologici e disciplinari.

Il portale è inoltre progettato per consentire l'interrogazione congiunta, dalla medesima interfaccia, dei cataloghi delle biblioteche del Sistema bibliotecario di Ateneo, delle riviste elettroniche sottoscritte ad abbonamento e dei prodotti della ricerca caricati su U-Gov, oltre che – naturalmente – delle diverse fonti digitali acquisite e catalogate/inventariate all'interno della SDL e delle risorse digitali di provenienza esterna, donate da utenti esterni o in quanto materiali di loro produzione o quali perché ne detengono i diritti d'uso.

All'interno della nostra comunità universitaria, come in molte altre istituzioni in ambito culturale e scientifico, esistono e vengono continuamente prodotti numerosi e importanti archivi e collezioni di diversa natura, ma senza dubbio rilevanti sul piano culturale e scientifico: un vasto universo documentario in forma analogica, digitalizzata o nativa digitale relativo a opere d'arte, monumenti e collezioni, archivi di ricerca, aree di scavo e siti preistorici, data set, musei virtuali (Capaldi, Ilardi, Ragone 2011; Ragone 2012). Solo alcuni di questi contenuti digitali sono accessibili sul web, in molti casi descritti in maniera piatta e non controllata e pubblicati su piattaforme diverse non integrate da un unico motore di ricerca. In altri casi, si tratta di materiale conservato in forma analogica, per lo più chiuso in armadi dimenticati e solo raramente, e non sempre correttamente, reso disponibile all'interno della stessa comunità disciplinare di cui il produttore o detentore del patrimonio pur è parte.

L'obiettivo prioritario di SDL è quello di integrare e rendere visibili attraverso un accesso unificato collezioni e archivi finora dimenticati o poco valorizzati, supportando strutture dell'Università e soggetti esterni che vogliono comunicare i propri patrimoni attraverso il portale SDL, in molti casi rivitalizzando mediante l'accessibilità sul web patrimoni informativi e di conoscenza specialistici sedimentati (ma non ordinati né valorizzati) nel corso di attività di ricerca, di cui tradizionalmente ci si limita a pubblicare i risultati principali. Le istituzioni del mondo culturale hanno un ruolo fondamentale nell'assicurare la rilevanza scientifica dei contenuti digitali e nel mediare tra i complessi documentari e gli utenti e garantiscono la qualità di tale processo anche mediante lo sviluppo di strumenti di ricerca appropriati in grado di salvaguardare il contesto documentario e di produzione e apparati descrittivi controllati, di qualità scientifica e filologicamente rigorosi.

La trasposizione in digitale non deve alterare la natura organica e sistemica degli archivi e delle collezioni, né impoverirne il contenuto informativo e le relazioni, ma deve piuttosto tendere a fornire qualità di contenuti e affidabilità. Il web offre nuovi strumenti di recupero e condivisione e consente di arricchire le risorse digitali o digitalizzate con informazioni di rappresentazione che descrivono il fitto tessuto di relazioni verticali e trasversali interne alle collezioni/archivi o di collegamento tra archivi e/o tra risorse.

Il sistema deve inoltre garantire la conservazione a lungo termine del patrimonio digitale.

4.2. Biblioteche, archivi o collezioni?

Digital library è un termine ambiguo, quasi totalmente assente nei progetti di matrice archivistica o museale: gallerie multimediali, teche, archivi digitali sono i termini alternativi utilizzati.

Numerosi sono i gruppi di ricerca e gli studi che analizzano e definiscono il concetto di *digital library* (Guerrini, Gambari, 2002; Noiret, 2011; Weston 2002; Salarelli, Tammaro, 2006; Santoro, 2006; ARL²

² Association of Research Libraries. URL= http://www.arl.org/resources/pubs/mm_proceedings/126mmappen2.shtml [ultima visita 18.2.2013].

DELOS³, Digital Library Federation⁴; IFLA⁵; IRCDL⁶, TPD⁷).

La maggior parte dei sistemi informativi che forniscono l'accesso a contenuti digitali culturali utilizzano il termine *digital library* nel suo senso più vasto:

Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.⁸

Al di là della terminologia adottata per definire i sistemi informativi di accesso ai patrimoni culturali, è di cruciale importanza restituire le informazioni su provenienza e contesto delle risorse, con la consapevolezza che è necessario fornire un accesso informativo integrato e controllato dagli specialisti relativo ai patrimoni digitali resi disponibili online. Per evitare di cadere nel rumore o nel silenzio informativo in fase di recupero dell'informazione, è necessaria una continua ricerca e sperimentazione per calibrare in sede di acquisizione delle risorse la specificità e il dettaglio scientifico della descrizione con i modelli adottati per la restituzione dei dati, che devono agevolare la facilità d'uso e la comprensione degli strumenti di ricerca.

Questi aspetti sono connessi alla qualità delle rappresentazioni e centrali nella fase di progettazione iniziale. Naturalmente, un certo margine di compromesso è inevitabile, ma esiste un livello minimo accettabile perché non si perda il senso del lavoro descrittivo e si sviluppino soluzioni meno appiattite di quelle offerte dai motori di ricerca oggi disponibili.

³ Network of Excellence on Digital Libraries. URL= <http://www.delos.info/>.

⁴ Digital Library Federation. URL= <http://diglib.org/> [ultima visita 18.2.2013].

⁵ International Federation of Library Associations and Institutions URL= <http://archive.ifla.org/VI/5/op/udtop8/udtop8.htm> [ultima visita 18.2.2013].

⁶ Italian Research Conference on Digital Libraries. URL= <http://ims.dei.unipd.it/websites/ircdl/home.html> [ultima visita 18.2.2013].

⁷ Theory and Practice of Digital Libraries. URL= <http://tpdl.eu/> [ultima visita 18.2.2013].

⁸ Digital Library Federation. URL= <http://diglib.org/about/dldefinition.htm> [ultima visita 18.2.2013].

Le soluzioni di *digital library* sono valutabili se si circoscrivono i modelli di riferimento e si definiscono categorie specifiche di patrimoni. I primi modelli e le prime definizioni elaborate alcuni anni fa hanno peraltro natura molto generale come risulta anche dal manifesto di DELOS che nel 2007 considerava la *digital library* "as a tool at the centre of intellectual activity having no logical, conceptual, physical, temporal, or personal borders or barriers on information", sottolineando la novità di un nuovo paradigma per l'accesso ai patrimoni digitali:

It has moved from a content-centric system that simply organizes and provides access to particular collections of data and information, to a person-centric system that aims to provide interesting, novel, personalized experiences to users. Its main role has moved from static storage and retrieval of information to facilitation of communication, collaboration, and other forms of interaction among scientists, researchers, or the general public on themes that are pertinent to the information stored in the Digital Library. Finally, it has moved from handling mostly centrally-located text to synthesizing distributed multimedia document collections, sensor data, mobile information, and pervasive computing services.⁹

Non sembrano trovare facilmente collocazione e soprattutto univocità di trattamento in questo contesto alcune categorie di materiali informativi, tra cui i patrimoni documentari sedimentati nello svolgimento delle attività di ricerca, in gran parte non organizzati né catalogati, spesso resi disponibili online attraverso processi di digitalizzazione legati ai progetti e alle indagini in corso e difficilmente riconducibili alle categorie conosciute e governate in ambito disciplinare. Se da un lato, in molti casi si tratta di archivi o di raccolte di natura biblioteconomica, in altri, sempre più frequenti, le esigenze di comunicazione online portano i ricercatori e i produttori di risorse a voler e a dover condividere ambiti, tipologie e contenuti diversificati creando le premesse per una significativa contaminazione delle regole descrittive e dei modelli di rappresentazione della conoscenza.

Nel caso dei materiali prodotti in ambito universitario, in

⁹ DELOS, *The Digital Library Manifesto* (2007) URL= http://www.delos.info/index.php?option=com_odl_open&handle=delos/2006_other_0081&version=1&Itemid=333 [ultima visita 18.2.2013].

particolare, non si tratta quasi mai né di archivi in senso stretto, né di raccolte documentarie coerenti: i materiali spesso sono sottoposti a selezioni finalizzate alla comunicazione, larga parte della documentazione amministrativa è conservata da strutture diverse, a volte dispersa, a volte presente in modo frammentario e comunque raramente oggetto di trattamento descrittivo e ancor meno di pubblicazione; nel caso di ricerche svolte in partecipazione con altri enti una parte della documentazione collegata è conservata in altre strutture e non è accessibile. È impossibile in questi casi ricorrere a standard descrittivi di un solo ambito disciplinare, sia pure quello archivistico che più di altri sembra prestarsi, per l'ampiezza dei termini di riferimento, a comprendere una variegata gamma di produzioni documentarie.

Il termine *collezione* alla fine utilizzato per necessità operative più che per convinzione scientifica nel progetto che qui si presenta è altrettanto ambiguo del concetto di *digital library*, come risulta chiaramente dalla definizione, anzi dalle definizioni proposte dalla Society of American archivists e riprese dalle *Rules for archival description* elaborate dagli archivisti canadesi:

"Collection n. ~ 1. A group of materials with some unifying characteristic. - 2. Materials assembled by a person, organization, or repository from a variety of sources; an artificial collection"¹⁰.

Spesso infatti non è una proprietà o caratteristica comune degli oggetti che li tiene insieme, bensì – come indica la seconda definizione del termine – una precisa volontà di assemblare e collezionare documenti, testimonianze e oggetti, altre volte invece una collezione può essere il residuo di un archivio andato disperso o frammentato presso istituzioni diverse.

An artificial accumulation of documents of any provenance brought together on the basis of some commons characteristic, e.g. way of acquisition, subject, language, medium, type of document, name of collector, which may be treated for descriptive purpose as unit under common title (Bureau of Canadian Archivists, 1990, 2008, Appendix D p.3).

¹⁰ URL= <http://www2.archivists.org/glossary/terms/c/collection> [ultima visita 02.03.2013].

Nel caso specifico molti patrimoni attualmente in fase di digitalizzazione e metadatazione per l'esposizione sul portale SDL sono addirittura collezioni nate a posteriori – se così si può dire – da archivi di ricerca, nel senso che si tratta di materiale riordinato e organizzato per la fruizione in un momento successivo alla sua produzione, selezionando una parte della documentazione, che nella maggior parte dei casi non comprende tutti i materiali che sono stati accumulati in fase di espletamento delle proprie attività e funzioni. Molto spesso infatti, come si è già ricordato, il vero e proprio archivio amministrativo e di ricerca di missioni archeologiche, etnologiche o di altro tipo è in carico alle strutture amministrative delle istituzioni e ai ricercatori che hanno partecipato a vario titolo alla Missione. Spesso si tratta di enti, istituzioni, persone di diversa natura, del tutto indipendenti e, in alcuni casi anche appartenenti ad altre nazioni.

Il tema meriterebbe di essere ulteriormente approfondito, così come il contesto attuale di gestione condivisa di patrimoni digitali di diversa provenienza richiederebbe una maggiore attenzione nell'utilizzo di una terminologia appropriata eventualmente anche aggiornata con rigore e creatività. Ma quello che è qui importante sottolineare, relativamente alle esigenze operative che il progetto SDL ha dovuto affrontare, è che tutte le considerazioni fin qui esposte avvalorano la necessità di gestire i processi di digitalizzazione e pubblicazione online dedicando alla qualità delle informazioni di rappresentazione e descrizione molta più attenzione di quanto non si sia fatto nel passato anche in iniziative internazionali di grande rilevanza come Europea. Si tratta in particolare di riconoscere che la contaminazione resa necessaria dai sistemi digitali di comunicazione non può tradursi nell'adozione di un solo modello descrittivo e dei suoi linguaggi di riferimento (Dublin Core? MODS? EAD/EAC?), ma richiede che per ciascuna area informativa e per ciascun dominio si rispetti la specificità funzionale e si adottino le soluzioni che garantiscano la miglior qualità possibile nella restituzione delle informazioni necessarie. La convergenza dei sistemi informativi, per molti versi resa inevitabile dalle strategie di evoluzione del web, non può comportare l'eliminazione delle specificità senza produrre conseguenze negative sulla qualità del trattamento e della restituzione delle fonti. Allo stesso tempo le comunità scientifiche e professionali cui compete la responsabilità della comunicazione delle memorie

digitali, non possono continuare a dividersi frammentando e indebolendo l'offerta informativa. Gli interventi di mediazione interdisciplinare sono quindi necessari e implicano un notevole grado di complessità, tanto più se non si vuole rinunciare da un lato agli obiettivi della interoperabilità tra sistemi e universi di conoscenza, dall'altro alla specializzazione delle informazioni.

L'esigenza di integrazione tra i diversi domini che si occupano di patrimoni culturali è ormai così sentita nelle comunità di riferimento che nel 2011 le associazioni nazionali operanti nei vari ambiti disciplinari del settore dei patrimoni culturali (AIB¹¹, ANAI¹² e ICOM Italia¹³), hanno dato vita a un coordinamento permanente (MAB¹⁴) per esplorare le prospettive di convergenza tra i mestieri e gli istituti in cui operano i professionisti degli archivi, delle biblioteche, dei musei:

Il Coordinamento si propone come luogo di elaborazione di proposte e di azioni che pongano gli operatori professionali dei beni culturali nella condizione di sviluppare esperienze di collaborazione, di confrontarsi sulle criticità dei singoli ambiti disciplinari e di promuovere soluzioni organizzative, normative, tecnico-scientifiche per gli istituti culturali e per il patrimonio culturale in genere. (Atto costitutivo MAB, 2012, p.1).

Nel caso specifico, con riferimento alle scelte operate nel progetto SDL, ad esempio, allorché si è trattato di definire i modi per documentare le informazioni di contesto delle 'collezioni', si è ritenuto di dover far ricorso agli standard sviluppati in ambito archivistico, sia per quanto riguarda le componenti informative relative al soggetto/ai soggetti che hanno prodotto i materiali e li hanno versati alla SDL, sia e ancor più allorché si tratta di dar conto della specificità di sedimentazione della collezione medesima, delle sue articolazioni, della sua struttura diversificata che prevede partizioni, serie, sezioni diverse. Si è infatti

¹¹ Associazione Italiana Biblioteche URL= <http://aib.it> [ultima visita 02.03.2013].

¹² Associazione Nazionale Archivistica Italiana URL= <http://anai.org/anai-cms/> [ultima visita 02.03.2013].

¹³ International Council of Museum - Comitato Nazionale Italiano URL= <http://icom-italia.org/> [ultima visita 02.03.2013].

¹⁴ Musei Archivi, Biblioteche - Professionisti del patrimonio culturale URL= <http://mab-italia.org/> [ultima visita 02.03.2013].

verificato nei fatti che gran parte delle raccolte documentarie prodotte in ambito scientifico richiedono per molti aspetti un trattamento descrittivo di natura archivistica capace di individuare non solo il livello di aggregazione di ciascuna entità, ma anche e soprattutto di definire e restituire in fase di catalogazione e di ricerca le informazioni sul contesto e sulla struttura, foriere di ulteriore arricchimento semantico.

In molti casi le collezioni sono inoltre complessi dinamici, che si incrementano periodicamente di nuove risorse, di inedite relazioni e nuovi vincoli: la struttura della collezione deve quindi rispecchiare la loro complessità originaria e potenziale ed essere aperta a nuove inevitabili acquisizioni nonché a ricongiungersi virtualmente con fondi conservati in altri luoghi ma con cui condivide legami originari.

4.2.1. Set di metadati, integrazione e interoperabilità

Il set di metadati adottato dal progetto è il MODS¹⁵, verso cui vengono mappati gli standard di descrizione di struttura e contenuto appartenenti a diversi domini (MARC¹⁶, RDA, Dublin Core¹⁷, ISAD, EAD¹⁸, EAC¹⁹, LIDO²⁰)

L'armonizzazione delle pratiche catalografiche viene realizzata mediante un'attività continua di studio e di confronto tra diverse tradizioni di dominio, e con specifici accordi con le altre istituzioni nazionali deputate alla conservazione e alla fruizione del patrimonio culturale: ICAR²¹, ICCU²², ICCD²³. Come già in parte sottolineato,

¹⁵ Metadata Object Description Schema URL= <http://loc.gov/standards/mods/> [ultima visita 18.2.2013].

¹⁶ Machine-Readable Cataloging URL= <http://loc.gov/marc/> [ultima visita 18.2.2013].

¹⁷ URL= <http://dublincore.org/> [ultima visita 18.2.2013].

¹⁸ Encoded Archival Description URL= <http://loc.gov/ead/> [ultima visita 18.2.2013].

¹⁹ Encoded Archival Context URL= <http://www2.archivists.org/standards/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf>. URL= <http://eac.staatsbibliothek-berlin.de/> [ultima visita 18.2.2013].

²⁰ URL= <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/> [ultima visita 18.2.2013].

²¹ Istituto Centrale per gli Archivi URL= <http://icar.beniculturali.it> [ultima visita 18.2.2013].

²² Istituto centrale per il catalogo unico URL= <http://iccu.sbn.it/> [ultima visita 18.2.2013].

²³ Istituto centrale per il catalogo e la documentazione URL= <http://iccd.beniculturali.it/> [ultima visita 18.2.2013].

infatti, la normalizzazione non deve essere confusa con l'omologazione, né l'integrazione con la convergenza. Va intesa piuttosto – almeno questa è stata la scelta operata nel progetto SDL – come la ricerca di un minimo comune denominatore che permetta l'interrogazione cumulativa di materiali eterogenei descritti in base ai differenti domini disciplinari, senza per questo impoverire le informazioni, che sono poi restituite per intero nell'esplorazione della risorsa.

Non sono naturalmente poche le aree che hanno richiesto e che ancora richiedono una riflessione multidisciplinare finalizzata ad adottare soluzioni convincenti e a risolvere ambiguità interpretative, a cominciare dalla chiara definizione di ciò che una *digital library* descrive, con riferimento ad esempio alla distinzione tra l'oggetto originario e la *copia digitale* realizzata attraverso il processo di digitalizzazione e destinata alla pubblicazione online.

4.2.2. Oggetto fisico e oggetto digitale

Gli aggregatori di contenuti digitali che non siano solo cataloghi on line o inventari di patrimoni posseduti hanno la necessità di distinguere l'oggetto fisico dall'oggetto digitale e indicare quali metadati si riferiscono alla risorsa originaria e quali alla sua rappresentazione digitale.

Il rigore scientifico riguardo l'oggetto della descrizione e il mantenimento della catena di conservazione possono realizzarsi solo dando conto in maniera puntuale dei metadati che si riferiscono alla risorsa analogica e quelli riferiti al surrogato digitale: il ProvidedCHO Provided Cultural Heritage Object rappresentato dalla Web Resource dell'EDM (Europeana Data Model)²⁴ adottato dal portale Europeana, l'oggetto fisico reale e la sua rappresentazione digitale.

Il portale Europeana²⁵, in quanto aggregatore di metadati, una sorta di catalogo di secondo livello, raccoglie i dati sui patrimoni di diverse istituzioni culturali sul territorio europeo, senza essere responsabile né delle risorse originarie né della conservazione a lungo termine dei materiali digitali. Ha potuto quindi sciogliere il nodo in questione con una certa facilità, anche se è giunto a definire con chiarezza il problema solo recentemente.

²⁴ URL= <http://europeana.eu/schemas/edm/> URL= <http://europeanconnect.eu/results-and-resources.php?page=1> [ultima visita 18.2.2013].

²⁵ URL= <http://europeana.eu/portal/> [ultima visita 18.2.2013].

Nel caso invece di una *digital library* che sia aggregatore ma anche produttore e collettore di archivi, collezioni e materiali – digitalizzati o nativi digitali - che forma nell'esercizio della sua attività di ricerca e di cui ha quindi la responsabilità scientifica e di conservazione digitale, i nodi teorici da affrontare sono ancor più complessi e non sempre facilmente distinguibili. La doppia funzione di ente produttore e depositario della catena di conservazione degli originali e di canale di comunicazione di patrimoni non propri e di cui non ha la responsabilità della produzione, peraltro posseduti se non in digitale, costituisce infatti un fattore di complessità a vari livelli: concettuali al momento della catalogazione, ma anche organizzativi in sede di definizione dei flussi di lavorazione e di responsabilità.

Nel nostro caso, SDL non è l'istituzione deputata alla conservazione e all'accesso alle collezioni fisiche: è un catalogo per così dire di 'secondo livello' e può rimandare al catalogo dell'istituzione custode dei relativi oggetti analogici o a un'interfaccia di interrogazione specifica per database complessi. In questo caso, peraltro, la catena di conservazione va preservata, e anche se l'accesso ai documenti originari non è regolato da SDL, è tuttavia indispensabile stabilire e rendere visibile il raccordo con il catalogo 'primario'. Tale raccordo avviene a più livelli: innanzitutto tramite la collocazione del documento fisico, il cui inserimento e aggiornamento va curato dalle strutture che forniscono/versano i materiali cui compete garantirne la disponibilità e l'accesso. In secondo luogo, tale raccordo si esprime mediante la descrizione, secondo standard archivistici desunti dal modello adottato dal Sistema archivistico nazionale e coerente con gli standard EAD/EAC, del soggetto produttore e del soggetto versante come si dirà in seguito.

Questa distinzione dei diversi piani concettuali permette inoltre di preservare le informazioni sulla provenienza e di aggregare materiali diversi dando conto nei metadati descrittivi e gestionali-amministrativi se i dati stessi si riferiscono all'oggetto fisico, alla sua rappresentazione digitale, o se invece le risorse sono native digitali.

Quindi i possibili oggetti della catalogazione sono:

- oggetti born digital
- rappresentazioni digitali di originali analogici, con link alla localizzazione della risorsa originaria.
- riproduzioni di oggetti che non possiedono un proprio catalogo o inventario o che non siano conservati in uno specifico museo o istituzione (ad esempio la riproduzione fotografica di un paesaggio o di uno scavo), che quindi non avranno il collegamento alla catalogazione dell'esemplare fisico originario.

In questa logica vengono sciolte le ambiguità riscontrate nei campi portatori di maggiore sovrapposizione semantica: autore o *creator*, data e luogo. Queste criticità sono particolarmente rilevanti nel caso di oggetti fisici di natura non documentaria ma di diverso tipo: reperti archeologici, opere d'arte, oggetti museali.

Il concetto di autore è inevitabilmente ambiguo nel contesto multidisciplinare in cui si opera, non solo per quanto concerne l'individuazione dell'agente (autore della rappresentazione fotografica o dell'oggetto rappresentato, ambito quest'ultimo che ricade senz'altro nella semantica, ma che spesso viene imprecisamente metadato come *creator*) ma anche nel caso di patrimoni di tipo archivistico o museale in cui può frequentemente non coincidere con l'autore dell'oggetto descritto, trattandosi piuttosto di autore nell'accezione di soggetto produttore dell'archivio o dell'entità che ha raccolto e organizzato una collezione di oggetti o rappresentazioni documentarie (analogiche o digitali) di attività di ricerca.

L'assunto di partenza è che va distinto nettamente l'ambito della risorsa originaria analogica (laddove esistente) e le relative competenze di conservazione e accesso ai documenti che sono in capo alle strutture detentrici; e il piano dell'oggetto digitale, che costituisce un livello di descrizione autonomo, esclusivamente semantico, da collegare all'eventuale descrizione dell'originale. Tale distinzione deve peraltro essere innanzitutto definita nella fase preliminare di descrizione della collezione e delle sue partizioni. È infatti in quella sede che è possibile chiarire con maggiore coerenza (anche con il supporto del responsabile scientifico della collezione) obiettivi e natura del progetto di digitalizzazione e stabilire ruoli e qualifiche degli attori che a vario titolo sono presenti nella produzione delle singole risorse, non diversamente del resto da quanto avviene nella descrizione di materiali d'archivio.

La costruzione di un solido impianto concettuale intorno all'oggetto della descrizione è del resto alla base dell'architettura dei contenuti e permette di creare un catalogo che in fase di ricerca restituisca risultati pertinenti e riduca il rumore informativo al minimo.

4.2.3. Descrizione a più livelli dei complessi documentari

Nel sistema progettato per la SDL, la descrizione delle collezioni – a differenza di quanto previsto nella gran parte delle *digital library* – avviene rispettando e rappresentando adeguatamente la struttura a più livelli del patrimonio documentario e le serie archivistiche, in armonia con gli standard di dominio e in considerazione del fatto che il concetto di collezione qui utilizzato è in molti casi, come si è già sottolineato, riconducibile a quello di archivio, anche se le aggregazioni archivistiche descritte non sono sempre quelle originarie.

È del resto necessario preservare il concetto di descrizione a più livelli per rappresentare patrimoni informativi/documentari portatori di strutture interne profonde e complesse e relazioni esterne molteplici.

Nello specifico il sistema di *digital library* gestisce con cura il sistema di relazioni che collegano le partizioni della collezione e i rapporti tra le collezioni e i diversi soggetti che ne sono all'origine o che si sono avvicendati nella loro gestione e custodia, riunendo concettualmente fondi che sono fisicamente dispersi (Yeo, 2012).

4.2.4. Descrizione dei soggetti produttori e versanti e documentazione di progetto

Se, come abbiamo visto, nel concetto di collezione è intrinseca l'intenzionalità nel riunire i materiali e sono necessarie delle scelte per selezionare e organizzare logicamente le risorse, in un progetto di digitalizzazione è di fondamentale importanza dare conto della storia e della provenienza della collezione e dichiarare i tipi di intervento e i criteri che li hanno guidati. Solo in questo modo è possibile rendere trasparenti questi processi e intellegibili le fonti documentarie e la storia culturale di cui sono portatori.

Gli apparati informativi di corredo si rendono indispensabili, oltre che per garantire ricchezza informativa e contestualizzazione dei patrimoni, anche per la certificazione della provenienza del docu-

mento e delle responsabilità in capo a singoli individui o istituzioni. Infatti "l'autenticità del documento può essere dimostrata o presunta se le informazioni di contesto amministrativo e documentario si sono mantenute e sono ricostruibili e ricercabili" [Guercio, 2010, p.28]. Questo è ancor più vero in ambiente digitale, dove è opportuno esplicitare ogni tipo di informazione fornendo una descrizione accurata della provenienza e della storia delle collezioni alle quali appartengono i singoli oggetti. La documentazione di contesto e di progetto è essenziale per garantire autorevolezza scientifica alla catena di conservazione e integrità e autenticità alle fonti documentarie: ambedue elementi necessari per stabilire un rapporto fiduciario tra i *service provider* e gli utenti.

In virtù di queste considerazioni particolare attenzione è stata posta nell'assicurare la descrizione dei soggetti produttori e di quelli versanti, che nel contesto della SDL acquistano peso giuridico oltre che un ruolo organizzativo. Nello specifico il sistema di *digital library* gestisce con cura il sistema di relazioni che collegano le partizioni della collezione e i rapporti tra le collezioni e i diversi soggetti che ne sono all'origine o che si sono avvicinati nella loro gestione e custodia.

Nel sistema SDL i progetti di digitalizzazione sono stilati in collaborazione tra gli esperti di dominio archivistico e bibliotecario e i responsabili scientifici e tecnici delle collezioni. Oltre a una descrizione dell'archivio/collezione effettuata secondo le norme del SAN²⁶ e nel rispetto degli standard ISAD(G)²⁷ /ISAAR(CPF)²⁸, ogni complesso documentario sottoposto a digitalizzazione, metadattazione e/o normalizzazione dei dati, è corredato da una scheda di progetto che fornisce informazioni sulla storia dell'archivio/collezione, sulle re-

²⁶ Sistema Archivistico Nazionale URL= <http://san.beniculturali.it> [ultima visita 18.2.2013].

²⁷ General International Standard Archival Description URL= <http://ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html> [ultima visita 18.2.2013].

²⁸ International Standard Archival Authority Record for Corporate Bodies, Persons and Families URL= <http://ica.org/10203/standards/isaar-cpf-international-standard-archival-authority-record-for-corporate-bodies-persons-and-families-2nd-edition.html> [ultima visita 18.2.2013].

sponsabilità tecniche, scientifiche ed editoriali coinvolte, sul soggetto produttore e sul soggetto versante, sulla consistenza della raccolta, sui criteri di selezione per la digitalizzazione e sulle fasi e gli incrementi previsti. Il progetto scientifico di digitalizzazione o di recupero di patrimoni digitali costituisce parte integrante della descrizione, una sorta di nota archivistica in grado di fornire elementi cruciali sulla storia del contesto tecnologico, istituzionale, amministrativo e sulla provenienza intesa anche come definizione di responsabilità specifiche per la produzione e gestione di patrimoni documentari digitali.

4.2.5. Normalizzazione del linguaggio: authority file e thesauri

Al fine di garantire interoperabilità e coerenza nella descrizione, la definizione e la gestione dei metadati nel progetto SDL ha implicato una notevole attenzione in termini di qualità dei processi di normalizzazione. Si è fatto perciò costante ricorso a authority file e thesauri la cui tipologia è stata individuata con riferimento agli standard internazionali, coerentemente con i diversi materiali descritti e nell'ottica dei linked data: TGN²⁹ e Geonames³⁰ per i luoghi, VIAF³¹ per i nomi di persona ed ente, PICO 4.3³² e Marc Code List for Relators³³ per i ruoli, PICO e il Nuovo soggettario di Firenze³⁴ per la soggettazione, con la possibile integrazione di vocabolari controllati disciplinari.

Grazie all'inserimento controllato dei nomi, dei descrittori, delle date e dei luoghi è possibile perciò creare dei punti di accesso realmente efficaci nell'indicizzazione e nella contestualizzazione spaziotemporale dei contenuti.

²⁹ Getty Thesaurus of Geographic Names URL= <http://getty.edu/research/tools/vocabularies/tgn/> [ultima visita 18.2.2013].

³⁰ Geonames URL= <http://geonames.org/> [ultima visita 18.2.2013].

³¹ Virtual International Authority File URL= <http://viaf.org/> [ultima visita 18.2.2013].

³² Portale Italiano della Cultura on line URL= http://culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml [ultima visita 18.2.2013].

³³ URL= <http://loc.gov/marc/relators/relaterm.html> [ultima visita 18.2.2013].

³⁴ URL= <http://thes.bncf.firenze.sbn.it/> [ultima visita 18.2.2013].

4.2.6. Catalogazione descrittiva, semantica e relazionale

La letteratura di settore, la ricerca e l'applicazione sperimentale di piattaforme, modelli di dati e ontologie per la realizzazione di sistemi integrati di accesso al patrimonio culturale e della ricerca stanno attraversando una fase di riflessione, grande fermento e transizione. Negli ultimi quindici anni sono stati realizzati numerosi progetti di digitalizzazione e/o aggregazione di archivi a ogni livello: locale, nazionale ed europeo, come nel caso di Europeana. Tuttavia, non si è stati ancora capaci di sfruttare appieno le potenzialità degli strumenti digitali. Nella gran parte dei progetti di *digital library* finora realizzati si è ad esempio rinunciato a mantenere e integrare le soluzioni descrittive delle diverse tradizioni disciplinari consolidate, privilegiando scelte 'parziali' che hanno in molti casi reso disponibili collezioni digitali dotate di limitate o del tutto inesistenti informazioni di contesto, senza indicazioni sulle aggregazioni e sulle relazioni interne, con descrizioni limitate alla sola collezione e ai singoli oggetti.

Un buon catalogo di biblioteca digitale dovrebbe invece esplicitare anche il tessuto di relazioni implicite fra i documenti che descrive ed espone, superando i limiti del pensiero lineare guthenberghiano e ripensando il catalogo come una rete di relazioni semantiche e spazio-temporali.

La Library of Congress ha avviato una profonda riflessione sull'opportunità di superare il concetto ormai riduttivo di descrizione elaborato in ambito bibliotecario e orientato all'esemplare e al formato fisico, per orientarsi verso una descrizione di tipo semantico e relazionale che tenga conto di diverse tipologie di materiale, superando i limiti che impone il formato Marc, nato quaranta anni fa in ambito bibliotecario e più volte rielaborato, ma ormai insufficiente per rappresentare materiali non bibliografici e relazioni complesse, come si legge nel report *On the record* della Library of Congress:

Different communities of bibliographic practice have grown up around different resource types: library collections of books and journals; archives; journal articles; and museum objects and images. As these resources and others become increasingly accessible through the Web, separation of the communities of practice that manage them is no longer desirable, sustainable, or functional. Bibliographic control is increasingly a matter of managing relationships — among works,

names, concepts, and object descriptions — across communities. Consistency of description within any single environment, such as the library catalog, is becoming less significant than the ability to make connections between environments [...].

(Library of Congress, 2008, p. 9).

La rivoluzione digitale ci permette e ci impone di superare una catalogazione che sia attenta alla semantica in maniera disomogenea (sia per adozione di strumenti di controllo terminologico che per applicazioni pratiche) per portarci verso una catalogazione focalizzata sulla semantica e sulle relazioni di contesto, semantiche e strutturali e sia in grado di normalizzare i vocabolari integrandoli al fine di assicurarne anche la specializzazione.

Lo scopo è ricondurre la conoscenza a uno stadio di ricchezza che possa servire a descriverne le varie forme e manifestazioni in modo univoco, consentendo la compresenza di materiali eterogenei senza perderne le informazioni di provenienza ed esplicitando le relazioni implicite tra le risorse.

La sfida che SDL ha voluto raccogliere risponde all'esigenza, da tempo dichiarata anche se scarsamente perseguita nei diversi progetti di biblioteche digitali, di assicurare un accesso alle risorse arricchito di informazioni descrittive grazie a nuove forme di intermediazione scientifica e fornite dagli specialisti di dominio, da coloro cioè che sono chiamati a "istituire, coordinare e connettere sintassi di segni" (Serrai, 1997, p. 29, nota 1).

4.3. Conclusioni

I sistemi informativi digitali permettono organizzare i materiali logicamente e semanticamente, restituendo i legami di contesto e la cornice di significato delle risorse. La normalizzazione della terminologia dei descrittori, l'utilizzo dei linguaggi e delle grammatiche del web semantico e la modellazione della realtà e degli oggetti per mezzo delle ontologie e degli standard di descrizione, possono coadiuvare l'ambizione che in fase di ricerca venga restituita un'informazione di qualità, controllata e non ridondante.

Il materiale grezzo acquista significato solo se inserito in una rete semantica e di riferimenti spazio-temporali, che a sua volta viene arricchita dalle singole risorse.

Una *digital library* non è semplicemente la riproposizione in forma elettronica, organizzata linearmente, del catalogo cartaceo, né un semplice deposito di risorse digitali, bensì è un complesso strumento di organizzazione della conoscenza. È necessario superare il concetto di catalogo elettronico come semplice trasposizione in digitale di collezioni fisiche, e sperimentare linguaggi, strumenti e formalismi atti a descrivere ed esplorare l'universo rappresentato fornendo all'utente, oltre alle informazioni sulle singole risorse e sulla loro semantica, le informazioni di contesto sull'ambiente di produzione collocato nelle coordinate spazio-temporali che gli sono proprie. È a questo complesso insieme di obiettivi che il progetto SDL si è ispirato e ha l'ambizione di contribuire.

4.4. Bibliografia

- ARMS WILLIAM Y. (2012) *The 1990s: The Formative Years of Digital Libraries* in «Library Hi Tech», vol 30, no. 4 URL= <http://www.emeraldinsight.com/journals.htm?articleid=17065297> [ultima visita 02.03.2013].
- BUREAU OF CANADIAN ARCHIVISTS (1990, revised 2008) *Rules for Archival Description*, Appendix D, Glossary URL= http://www.cdncouncilarchives.ca/RAD/RADComplete_July2008.pdf [ultima visita 02.03.2013].
- CALVO M., CIOTTI F., RONCAGLIA G., ZELA MARCO A. (2004). *Internet 2004*, Laterza. URL= <http://www.marcocalvo.it/libri/internet/home/index.htm> [ultima visita 02.03.2013].
- CAPALDI D., ILARDI E., RAGONE G. (2011). *I cantieri della memoria*, Liguori.
- CURRAL J., MOSS M., STUART S. (2004) *What is a collection?* «Archivaria», no 58 URL= <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12480/13594> [ultima visita 02.03.2013].
- DELOS, *The Digital Library Manifesto* (2007) URL= http://www.delos.info/index.php?option=com_odl_open&handle=delos/2006_other_0081&version=1&Itemid=333 [ultima visita 02.03.2013].
- LIBRARY OF CONGRESS, *Bibliographic Framework Transition Initiative*. URL= <http://loc.gov/marc/transition> [ultima visita 02.03.2013].

- LIBRARY OF CONGRESS (2008). *On the record report of the Library of Congress Working Group on the Future of Bibliographic Control*. URL= <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf> [ultima visita 02.03.2013].
- YEO G. (2012). *The Conceptual Fonds and the Physical Collection*. «Archivaria», no. 73 URL= <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/13384/14691> [ultima visita 02.03.2013].
- GUERCIO M. (2010). *Archivistica informatica*, Carocci.
- GUERRINI M., GAMBARI S. (2002). *Definire e catalogare le risorse elettroniche* Editrice Bibliografica. URL= <http://eprints.unifi.it/archive/00001949/> [ultima visita 02.03.2013].
- MAB (2012). *Atto costitutivo* URL= http://www.mab-italia.org/images/pdf/MAB-Naz_Atto-costitutivo_2012-06-12.pdf
- METITIERI F., RIDI R. (2007). *Biblioteche in rete*, Laterza. URL= <http://laterza.it/bibliotecheinrete/> [ultima visita 02.03.2013].
- NOIRET S. (2011). *Storia digitale* in Perilli L., Fiormonte D., a c. di, *La macchina nel tempo*, Le Lettere, pp. 201-258.
- RAGONE (2012). *Introduzione ai lavori*. Quaderni DigiLab», no 2 URL= http://digilab-epub.uniroma1.it/index.php/Quaderni_DigiLab/article/view/17/15 [ultima visita 02.03.2013].
- SALARELLI A., TAMMARO A. (2006). *La biblioteca digitale*, Editrice Bibliografica.
- SANTORO M. (2006). *Biblioteche e innovazione*, Editrice Bibliografica.
- SERRAI A. (1997). *Introduzione*. In: *Storia della bibliografia*, Bulzoni, vol.7.
- WESTON P.G. (2001). *Catalogazione bibliografica: dal formato MARC a FRBR*. «Bollettino AIB», vol 41, no 3 URL= <http://bollettino.aib.it/article/view/8536/7616#34> [ultima visita 02.03.2013].
- WESTON P.G. (2002). *Il catalogo elettronico*, Carocci.
- VITALI S. (2004). *Passato digitale*, Bruno Mondadori.

PARTE II

ARCHIVI DIGITALI; ARCHITETTURE E INTERFACCE

5. Il ruolo dell'*Information Visualization* nella progettazione di interfacce per archivi digitali eterogenei

*Michele Mauri*¹, *Paolo Ciuccarelli*²

Abstract. L'utilizzo di metodologie provenienti dalla information visualization nella progettazione di interfacce a collezioni digitali è una pratica che si sta consolidando negli ultimi anni (Hilton, 2010), con l'obiettivo di superare l'accesso puntuale ai singoli elementi a vantaggio di una visione complessiva dei contenuti. Allo stato attuale queste metodologie sono principalmente utilizzate per costruire singole viste sulle collezioni, concentrandosi su un'unica tipologia di elementi e progettando un'interfaccia che ne rispecchi le caratteristiche. Nel caso studio presentato, l'archivio Baldessari, si mostrano le potenzialità della information visualization non solo per costruire singole viste, ma anche e soprattutto nella progettazione di un linguaggio visuale coerente che permetta la fruizione unitaria di elementi tra loro eterogenei.

Parole chiave: information visualization, digital libraries, browsing interfaces, visual language.

¹ Politecnico di Milano, Milano, Italia. e-mail: michele.mauri@mail.polimi.it.

² Politecnico di Milano, Milano, Italia. e-mail: paolo.ciuccarelli@polimi.it.

5.1. Introduzione

Nella progettazione di interfacce per collezioni e archivi digitali, le modalità di accesso ed esplorazione visuale delle informazioni stanno diventando un elemento sempre più rilevante. A fronte del processo di digitalizzazione degli ultimi anni, unito alle potenzialità delle tecnologie emergenti come il *Semantic Web*, diventa sempre più urgente trovare modalità di accesso alle informazioni capaci di restituirne la ricchezza agli utenti (Cortese, 2011).

Più volte la letteratura sull'argomento (Schreibman, 2004) ha evidenziato come la potenzialità delle collezioni digitali non sta nel rendere più veloce o precisa l'esecuzione di operazioni effettuate su collezioni analogiche, ma nell'abilitare azioni completamente nuove, superando i vincoli fisici e offrendo nuove prospettive sulla fruizione dei contenuti. La traduzione digitale dei metadati permette di operare su di essi operazioni tipiche dei dati digitali, come aggregazioni e filtraggi. Ciò pone le basi per il superamento dell'accesso puntuale alle informazioni a favore di una fruizione in cui l'utente possa avere una panoramica generale del contenuto dell'archivio per poi entrare in contatto con il singolo elemento. Questo cambio di accezione ha importanti ripercussioni sulla progettazione dell'accesso alle informazioni, suggerendo un uso del metadato come mezzo di aggregazione, raggruppamento e filtraggio.

Mitchell Whitelaw e Sam Hinton paragonano il classico campo di ricerca, principale strumento di esplorazione delle collezioni digitali, a un accesso "fisico" alle informazioni (Hinton 2010). A seconda del tipo di utenza il campo di ricerca può diventare un muro che separa l'utente dal contenuto della collezione stessa, in quanto non ha modo di conoscerne l'estensione, la ricchezza, le caratteristiche. Una persona esperta del dominio conosce già il contenuto della collezione, e necessita uno strumento che lo colleghi velocemente all'elemento che sta cercando: il campo di ricerca è uno strumento efficiente e rapido per ottenere questo risultato. Al contrario, un utente generico non sempre ha un obiettivo preciso (Marchionini 2006): in questo caso il campo di ricerca non aiuta l'utente, il quale può identificare i contenuti per lui rilevanti solo con un processo di prove ed errori. Infine l'utente

talvolta non ha interesse nel singolo elemento, quanto ad avere una più ampia panoramica sui contenuti dell'archivio: ad esempio quali sono le categorie con più elementi, o la loro distribuzione temporale. Il campo di ricerca è un esempio di "interfaccia avara", cui si contrappone il concetto di "interfaccia generosa" (Whitelaw, 2012). Due tipologie di interfaccia che presuppongono, secondo Whitelaw, un *ethos* differente: la seconda in particolare deve fornire all'utente il maggior numero di informazioni possibili e offrire strumenti per raffinare le informazioni visualizzate.

I due tipi di accesso sono tra loro complementari, ed hanno profonde implicazioni sulla *user experience*, modificando radicalmente le possibilità di azione dell'utente. Il classico campo di ricerca è un accesso di tipo *additivo* ai contenuti: l'utente parte con uno spazio bianco che popola in base alla propria ricerca. Alla sua domanda, codificata in una *query* testuale, corrisponde un set di risultati. In questo caso l'utente ha un accesso diretto ai singoli elementi, ma senza avere informazioni sul loro contesto: non conosce la dimensione dell'archivio in cui esegue la ricerca, quali sono le chiavi di ricerca utilizzabili, quali tipologie di oggetti contiene l'archivio. Una modalità di tipo esplorativo, utilizzando un'interfaccia "generosa", offre all'utente un percorso di tipo *sottrattivo*: all'utente vengono presentati tutti gli elementi presenti nell'archivio (*overview*) offrendo strumenti per filtrarli fino ad arrivare al singolo elemento. In questo caso l'utente ha sia la possibilità di avere informazioni sul contesto in cui svolge la ricerca, sia un accesso diretto ai singoli elementi. Le due modalità di accesso non sono tra loro esclusive, ed una soluzione ibrida, in cui l'utente può scegliere quale utilizzare, è auspicabile.

La creazione di interfacce "generose" è ancora ad uno stadio di sperimentazione, ed è un campo promettente per la *information visualization*. Costruire questo tipo di interfacce porta a confrontarsi con problematiche sia tecnologico/computazionali (di volumi di informazione da gestire) che grafico/comunicative: come "mostrare tutto" all'utente senza creare confusione? La necessità di trovare nuove forme di rappresentazione diventa prioritaria nella costruzione delle interfacce.

5.2. Interfacce e *Information Visualization*

La sfida nella progettazione di questo tipo di interfacce non si esaurisce nel mostrare più informazioni possibili all'utente, ma nel presentarle in modo che siano generatrici di conoscenza. La *information visualization*, scienza che si occupa della rappresentazione visuale dei dati (Friendly, 2009), ha sviluppato metodologie utili a risolvere le problematiche poste da questo nuovo approccio alla creazione di interfacce per collezioni digitali.³

5.2.1. Viste sui dati

Progettare una interfaccia capace di "mostrare tutto" rispettando i vincoli tecnici e visivi posti dalla fruizione attraverso uno schermo pone la necessità di identificare delle modalità di aggregazione dei dati visualizzati. Il problema non è di quantità dei dati, ma di qualità delle informazioni fornite a diversi livelli di aggregazione.

Come in una mappa cartografica abbiamo informazioni diverse a differenti scale, così nella progettazione delle interfacce è necessario identificare quali informazioni l'utente può ottenere man mano che passa da una visione generale a una puntuale. È ciò che Ben Shneiderman nel campo della *data visualization* definisce il "mantra della ricerca visuale": "*Overview first, zoom and filter, then details-on-demand*" (Shneiderman, 1996). L'idea del mantra è quella di riproporre in modo iterativo questo tipo di approccio con le informazioni, in modo da permettere all'utente di passare da una visione globale a quella puntuale.

Whitelaw, nel proporre la sua idea di interfaccia generosa, riprende e sviluppa questo concetto identificando cinque buone pratiche nella progettazione di interfacce:

Mostrare, non chiedere. L'autore evidenzia come, alla presenza del solo campo di ricerca, l'utente debba formulare una query pur non avendo un obiettivo preciso nella ricerca. L'utente dovrebbe invece essere accolto dai contenuti della collezione.

³ Per una panoramica sulla *information visualization* si consiglia la lettura della pubblicazione "What is Visualization?" di Lev Manovich (Manovich 2011). Per avere una idea completa della nascita e sviluppo della disciplina si rimanda al lavoro di Michael Friendly (Friendly 2009).

Offrire ricche visioni d'insieme. L'accesso dovrebbe essere visualmente generoso, offrendo elementi che aiutino l'utente a orientarsi e a comprendere le caratteristiche della collezione.

Fornire campioni. Mostrare l'intera collezione nella maggior parte dei casi è impossibile, e diviene importante fornire campioni visuali rappresentativi della collezione.

Fornire contesto. Le interfacce devono aiutare l'utente a esplorare la collezione, esplicitando i legami tra i diversi elementi conservati.

Offrire risultati di alta qualità. Molte delle informazioni su un artefatto non possono essere codificate nei metadati, perché troppo qualitative o soggettive. L'accesso a una versione ad alta qualità dei contenuti permette di convogliare queste informazioni che andrebbero altrimenti perse.

Sviluppando le interfacce per il progetto Baldessari i punti sopra elencati sono stati ripresi e sviluppati, testando potenzialità e limiti della *information visualization* in questo campo.

5.2.2. Viste multiple

Una collezione può essere composta di centinaia, migliaia di singoli elementi, ed ad ognuno di essi sono legati molteplici metadati (proprietà). Cosa significa 'mostrare tutto' in questo contesto? Come è possibile offrire una visione globale e coerente di tutto il contenuto della collezione? Mantenendo l'obiettivo di offrire una visione globale della collezione, diventa necessario definire su quali variabili concentrarsi, siano la dimensione temporale, quella geografica, o le similarità tra gli elementi.

Non essendo possibile sintetizzare tutte le informazioni disponibili in una singola rappresentazione visuale, diviene necessario progettare più viste sulla collezione, ognuna delle quali offra una differente prospettiva sui dati contenuti nella collezione. All'attuale stato dell'arte, in molte soluzioni la *information visualization* è stata utilizzata costruendo un'unica vista sulla collezione scelta (Ruecker 2011). L'innovazione del progetto "Archivio Baldessari" consiste nel presentare un'interfaccia costituita da più viste, o prospettive, sulla collezione, e nell'identificazione di modalità d'interazione che permettano all'utente di passare da una vista all'altra in modo coerente.

In questo progetto le metodologie della *information visualization* non sono impiegate solo per costruire una singola rappresentazione, ma

per progettare un linguaggio visuale adattabile a diverse tipologie di elementi ed in grado di mostrarne differenti proprietà. La *information visualization* viene adottata per identificare dei *pattern* grafico visuale e delle modalità di interazione in grado di guidare ed aiutare l'utente nel processo esplorativo.

5.3. L'archivio Baldessari

Il progetto presentato nasce dalla collaborazione tra due laboratori del Politecnico di Milano: LADA (Laboratorio Archivi di Design e Architettura) e DensityDesign Research Lab. Il Laboratorio LADA si occupa da più di venti anni di archivi di design e architettura, gestendo in particolare gli archivi Luciano Baldessari, Agnoldomenico Pica, Carlo Perogalli e la collezione Silvano Zorzi. Il laboratorio DensityDesign Research Lab si occupa di *information visualization* e rappresentazione visuale di fenomeni complessi. Oltre alle attività di ricerca, il laboratorio conduce attività didattica e consulenza scientifica per tesi magistrali e di dottorato.

L'occasione per una sperimentazione congiunta nasce dalla completa digitalizzazione dell'archivio Baldessari, e dalla collaborazione e con altri due archivi che, con il Politecnico, coprono tutta la produzione dell'architetto. Durante il progetto è stata progettata e parzialmente sviluppata una soluzione di gestione, arricchimento e divulgazione dell'archivio. La prima fase si è concentrata sulla progettazione dell'architettura delle informazioni necessaria a rendere esplorabile l'archivio. Le informazioni sono state poi integrate con opere provenienti dalle collezioni del MART (Museo di Arte Moderna e Contemporanea di Trento e Rovereto) e del CASVA (Centro di Alti Studi sulle Arti Visive del Comune di Milano). A partire da questo corpo di opere, sono state progettate le viste sull'archivio, identificando specifiche modalità di arricchimento delle informazioni. La realizzazione del progetto ha portato a una collaborazione con le persone che si occupano dell'archivio, permettendo di comprendere quali sono gli effettivi passaggi e le criticità nella realizzazione di interfacce per archivi.

Il laboratorio di ricerca DensityDesign porta in questo progetto sia le attuali competenze di Design della Comunicazione e *information visualization*, sia l'esperienza maturata tra il 2000 e il 2006, con il

progetto DesignNet, nella progettazione e sviluppo di sistemi per la gestione di archivi eterogenei a supporto dell'attività conoscitiva di ricercatori e studenti di design (Ciuccarelli, 2006).

La partecipazione del laboratorio fin dalla fase di strutturazione del database ha permesso di impostare un processo nel quale l'architettura delle informazioni è funzionale alla loro fruizione. La progettazione dell'interfaccia non è più conseguente a quella del database ma parte integrante, evitando situazioni in cui la fruizione delle informazioni è determinata dalla struttura delle informazioni stesse.

5.3.1. Struttura dell'archivio

La produzione dell'architetto Luciano Baldessari è suddivisa in tre principali archivi: l'archivio del Politecnico di Milano contiene disegni tecnici e carteggi, il CASVA (Centro di Alti Studi sulle Arti Visive) conserva la produzione artistica, e infine l'archivio del MART (museo di arte moderna e contemporanea di Trento e Rovereto) contiene la corrispondenza privata e le fotografie. Il progetto nasce con l'obiettivo di costruire una prima piattaforma sperimentale per un accesso remoto, via web, ai contenuti degli archivi.

L'eterogeneità delle fonti è un elemento caratterizzante di questo progetto; la prima fase del lavoro si è concentrata sull'identificazione di un modello di dati capace di accogliere e rispettare questa eterogeneità dei materiali. I diversi archivi presentano più tipologie di contenuto: carteggi, disegni, disegni tecnici, fotografie. Ognuno di questi elementi ha dei metadati che ne rispecchiano le caratteristiche, diversi tra di loro. Inoltre, anche in presenza della stessa tipologia di documento, ogni archivio ha una sua precisa modalità di archiviazione e descrizione. Piuttosto che definire uno schema a priori in cui forzare gli elementi provenienti dai tre differenti archivi si è optato per identificare una struttura in grado di accogliere e valorizzare l'eterogeneità dei dati. Allineare i metadati delle tre collezioni, infatti, avrebbe richiesto un lavoro preliminare non coerente con i tempi del progetto. Il compromesso è l'adozione della logica "*deliver in beta*", ovvero popolare un primo database con le informazioni disponibili, mantenendo eventuali buchi ed inconsistenze, e fornendo ai gestori delle collezioni strumenti per modificare, pulire ed allineare i dati contenuti. In questo modo la progettazione delle interfacce non si basa su un database monolitico e finito, ma segue e mostra anche l'evoluzione dei dati contenuti.

Per la prima parte del progetto ci si è concentrati su cinquanta progetti rappresentativi dell'architetto, con l'obiettivo di costruire una prima piattaforma tecnologica in cui potranno poi essere inseriti altri progetti in futuro.

In questa pubblicazione si descrive la logica con cui è stata creata l'architettura dei dati e la progettazione delle interfacce, mostrando come questi due momenti siano strettamente connessi tra di loro.

5.3.2. Structurare il database

Per confrontarsi con un pubblico generico, non costituito da soli esperti e studiosi, è fondamentale identificare quali sono gli elementi che compongono la collezione. Per elementi non intendiamo solo le tipologie dei documenti presenti, ma tutte le entità su cui l'utente può eseguire la propria ricerca: entità specifiche del dominio della collezione. Nel caso del progetto presentato, il dominio è legato all'architettura. Le entità principali sono i progetti architettonici realizzati dall'architetto Baldessari. A ogni progetto sono legate le diverse tipologie di artefatti disponibili, tra cui carteggi, fotografie, disegni tecnici, pitture. Ogni progetto ha coinvolto una rete di persone in qualità di collaboratori, co-progettisti, corrispondenti. Ogni progetto ha una precisa posizione geografica, in termini di stato, città, quartiere, via. Infine ogni progetto ha avuto una copertura bibliografica su pubblicazioni, riviste specializzate, quotidiani.

5.3.3. Tipologie di lettura

La letteratura sull'argomento (Schreibman, 2004) evidenzia come la potenzialità delle collezioni digitali non sta nel rendere più veloce o precisa l'esecuzione di operazioni effettuate su collezioni analogiche, ma nell'abilitare azioni completamente nuove superando i vincoli fisici ed offrendo nuove prospettive sulla fruizione dei contenuti. La progettazione di viste sulla collezione comporta una scelta del tipo di lettura che si vuole proporre all'utente. Non esistono soluzioni predefinite applicabili a qualunque collezione, ed è necessario confrontarsi con esperti del dominio per comprendere quali siano le più rilevanti.

Nel progetto è stato avviato un dialogo con i gestori degli archivi in modo da riconoscere quali tipi di lettura proporre all'utente. Quelle identificate sono:

- Temporale: quale è stato il periodo più proficuo per l'architetto? Quali progetti sono stati svolti in concomitanza? È possibile identificare un'evoluzione stilistica dell'architetto?
- Geografica: che tipo di distribuzione geografica hanno i progetti? Come si relaziona temporalmente?
- Relazionale: quali sono le persone che hanno collaborato con l'architetto? Come sono cambiate queste relazioni nel tempo? Quali sono i progetti maggiormente citati?

La struttura dell'archivio del Politecnico, il più corposo dei tre considerati, è basata sui progetti, cui sono associati artefatti, persone e bibliografia in forma di metadati. Ciò significa che prima dell'avvio del progetto non era possibile sapere, se non con un'accurata analisi delle fonti, quale fosse la persona con cui l'architetto ha maggiormente collaborato, né la distribuzione temporale dei progetti e della produzione artistica.

Il primo passaggio per la progettazione della piattaforma è stato l'estrazione delle entità individuate in modo da poterle strutturare in un database di tipo relazionale. Questa operazione è stata svolta utilizzando un processo semi-automatizzato in collaborazione con i gestori dell'archivio. La documentazione sui progetti è stata prodotta manualmente nel corso di anni, grazie alla collaborazione di più persone. Utilizzando degli script costruiti ad-hoc sono state estratte tutte le informazioni necessarie. Il primo popolamento del database presentava numerose inconsistenze, dovute ad errori di battitura e diverse modalità di annotazione (ad esempio, la stessa pubblicazione appariva in più di un progetto adottando diversi stili di citazione). Tali inconsistenze sono impossibili da risolvere in modo automatico, ed è necessario l'intervento di esperti del dominio per risolverle. I dati prodotti automaticamente sono quindi stati validati utilizzando una metodologia semi-automatizzata, in cui stringhe simili sono state raggruppate in modo automatico, e poi validate e corrette da studiosi.

Il database è poi stato arricchito appoggiandosi a basi dati esterne, aggiungendo nuovi metadati. Il livello geografico è stato costruito utilizzando il servizio di geo localizzazione fornito da Google Maps: utilizzando l'indirizzo associato a ogni progetto ne sono state estratte le coordinate geografiche.

5.3.4. Progettare l'interfaccia

Obiettivo generale del progetto è l'abilitazione di un approccio esplorativo ai dati. Ognuno degli elementi identificati è rappresentato utilizzando più modelli visivi, mantenendo una coerenza logica ed estetica tra di essi. Lo schema dell'interazione prevede i seguenti passaggi:

- **Definizione delle entità da visualizzare.** L'utente definisce quali sono le entità per lui rilevanti, scegliendo tra progetti, documenti, attori.
- **Definizione della vista.** L'utente sceglie da quale prospettiva esplorare i dati. A seconda della tipologia di entità scelta, saranno disponibili una o più viste.
- **Definizione dei filtri.** L'utente può applicare uno o più filtri per ridurre gli elementi visualizzati.

Questa architettura dell'interfaccia rispecchia i cinque punti fondamentali identificati da Whitelaw: mostrare, non chiedere; offrire ricche visioni d'insieme; fornire campioni; fornire contesto; offrire risultati di alta qualità. A questi viene aggiunto un ulteriore legame tra le viste progettate: le singole visualizzazioni non sono tra loro separate, e l'utente può agevolmente cambiare tipo di vista mantenendo gli stessi filtri. In questo modo l'utente può esplorare l'archivio per passaggi successivi, settando una serie di filtri; può cambiare vista, modificare nuovamente i filtri, e così via. Il processo iterativo di prove e raffinazioni successive, che alla presenza del solo campo di ricerca può diventare frustrante per l'utente, diviene invece efficace modo di esplorazione con una interfaccia di tipo *additivo*, in quanto l'utente ha possibilità di cambiare scala e dettaglio degli elementi visualizzati a seconda delle sue necessità.

5.3.5. Progettare le viste

La scelta delle viste è frutto di un confronto con i gestori dell'archivio, e i modelli visuali selezionati sono un compromesso tra capacità di offrire una visione di insieme e il possibile spaesamento dell'utente. I modelli adottati non sono quindi totalmente sperimentali ma presi da un bacino di soluzioni note e già testate come efficienti per la visualizzazione dei dati.

Le viste implementate nel progetto sono cinque, e variano dal livello più analitico e puntuale a uno più generale.

Lista. Vista analitica, permette di visualizzare una lista in formato tabulare di tutti i contenuti. Ogni riga rappresenta un elemento del database, di cui vengono restituiti i metadati salienti. La lista è riordinabile utilizzando uno qualsiasi dei campi riportati. La lista non è divisa in pagine ma scorribile all'infinito, caricando dinamicamente gli elementi man mano che l'utente si muove tra di essi.

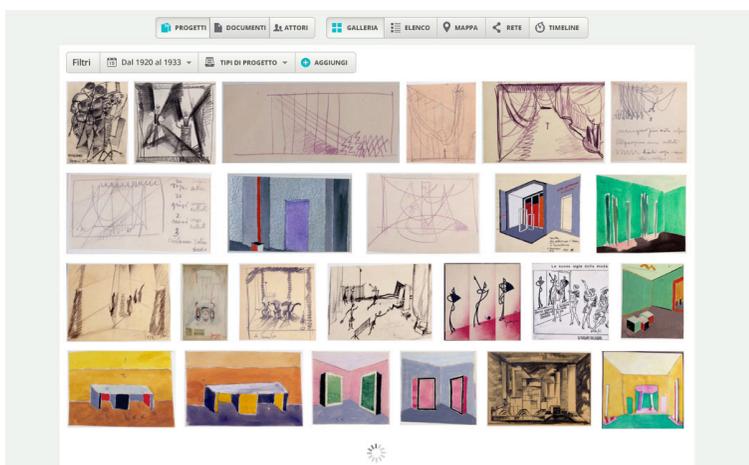


Fig. 5.1. Vista "thumbnails": ogni elemento viene visualizzato utilizzando una anteprima.

Thumbnails. Questa vista ripropone una visione dei singoli elementi in modo più compatto, non in forma tabulare, esaltando la natura figurativa dei contenuti. Mostrare i singoli elementi che compongono la collezione è importante: ci sono informazioni nel disegno tecnico, nella pittura e nella fotografia che non sono descrivibili attraverso il solo uso di metadati testuali. Avere accesso diretto agli artefatti permette letture di tipo stilistico, o di contenuto, altrimenti impossibili. Anche in questo caso si è preferito adottare una soluzione con scorrimento infinito rispetto ad una suddivisione in pagine dei risultati, per non interrompere l'esplorazione dell'utente.

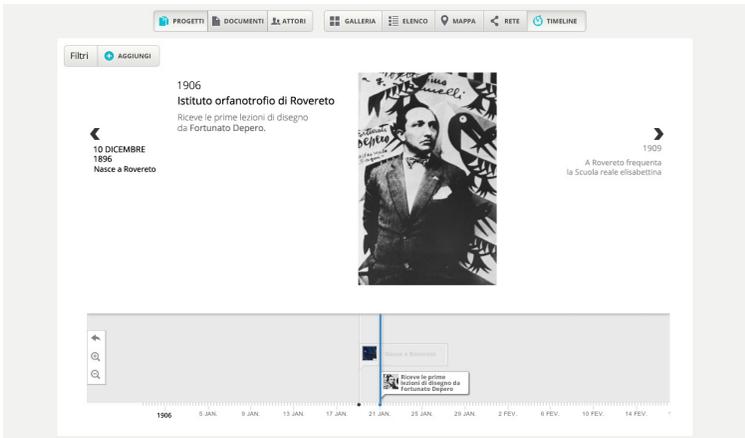


Fig. 5.2. Vista *timeline*: ogni evento temporale è posizionato su una linea temporale interattiva su cui l'utente può eseguire operazioni di zoom. Nella parte superiore ogni elemento è mostrato singolarmente in dettaglio.

Timeline. Questa vista dispone gli elementi temporalmente, utilizzando il modello visuale della linea temporale. In questa vista è possibile avere differenti livelli di dettaglio: è possibile scegliere il livello di scala temporale da utilizzare, così da poter passare da una visione generale su tutta la produzione dell'architetto ad una più puntuale su un preciso periodo storico. La *timeline* è divisa in due parti: in quella sottostante si ha una visione aggregata temporalmente, mentre in quella superiore è possibile navigare i singoli elementi uno ad uno.

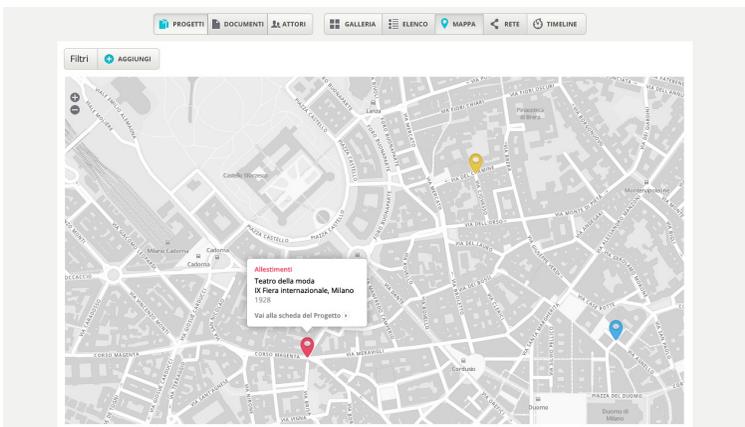


Fig. 5.3. Vista "mappa": gli elementi vengono posizionati ed aggregati geograficamente.

Mappa. Anche in questa vista è possibile aggregare le informazioni in modo da ottenere una immagine globale dei contenuti. Utilizzando i dati geografici estratti automaticamente dalle descrizioni dei progetti è stato possibile disporli su una mappa. A diversi livelli di scala, i progetti sono aggregati in modo da evitare sovrapposizioni e da restituire l'effettiva distribuzione spaziale. Quando uno o più progetti si trovano vicini vengono visivamente raggruppati con una circonferenza contenente il numero di risultati.

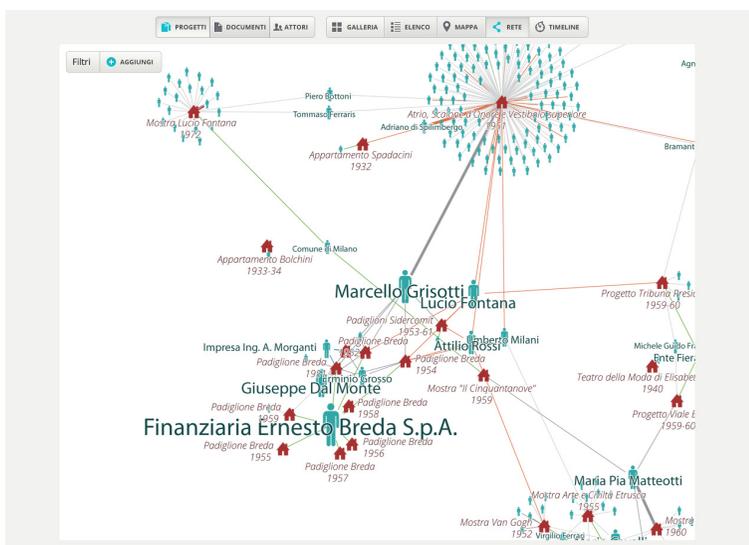


Fig. 5.4. Vista "rete": persone e progetti vengono visualizzati sotto forma di grafo a rete. L'utilizzo di icone e colori aiuta l'utente ad identificare il tipo di entità visualizzate.

Rete. Con questa vista è possibile esplorare la rete che lega persone (costruttori, coprogettisti, committenti) e progetti dell'architetto Baldessari. Peculiarità di questa rete è quella di non essere composta da una sola tipologia di elementi ma da due (persone e progetti). Questo tipo di rete è definito, nella teoria dei grafi, rete bipartita. Per enfatizzare questa caratteristica sono stati utilizzati due glifi differenti per identificare le entità. La struttura a rete permette di calcolare statistiche proprie della rete dei grafi, e ad ogni persona è stato associato un peso in base al numero di connessioni. Questo valore è rispecchiato dalla dimensione dei nodi. Anche le connessioni sono pesate in base al numero di relazioni intercorrenti tra persona e progetto. Nella visualizzazione di reti, una problematica è legata alla sovrappo-

sizione delle etichette: non essendo la posizione dei nodi definita a priori, ma conseguente all'intreccio delle relazioni, capita che più etichette si sovrappongano. La soluzione adottata è mantenere solo l'etichetta del nodo con il più alto numero di connessioni. Interagendo con la rete l'utente può comunque avere il nome anche dei nodi più piccoli. La rete è esplorabile selezionando una persona o un progetto. In base alla selezione, saranno mostrate le connessioni fino a due livelli di distanza.

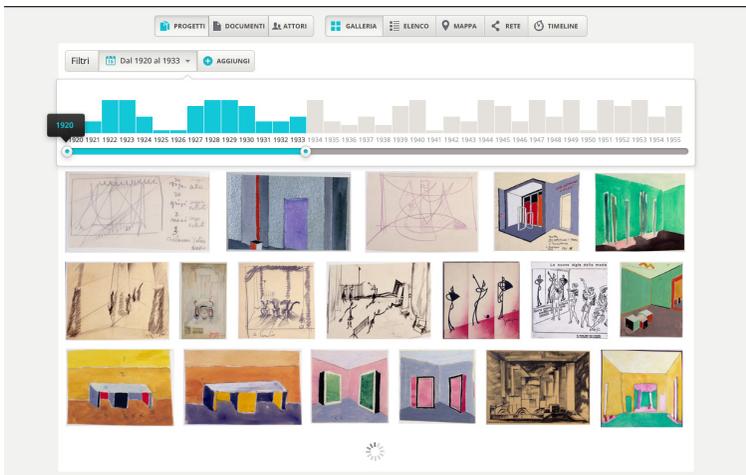


Fig. 5.5. Esempio di filtro visuale. Quando si aggiunge un filtro temporale, appare un pannello contenente un istogramma che permette di vedere il numero di elementi presenti per ogni anno.

Gestione dei filtri. In ogni vista è possibile filtrare i risultati in base alle proprietà degli elementi visualizzati. I filtri permangono quando si cambia vista, e l'utente ha la possibilità di vedere lo stesso set di elementi con diverse modalità di aggregazione. I filtri sono concatenabili secondo una logica 'AND': è possibile, ad esempio filtrare i progetti appartenenti ad un determinato periodo storico e ad una particolare categoria. Anche per l'interfaccia dei filtri, ove sensato, è stata utilizzata la visualizzazione dei dati. Il pannello per filtrare temporalmente, ad esempio, è visualizzato sotto forma di istogramma, in modo da comunicare la densità di elementi per ogni anno.

5.4. Futuri sviluppi

Strutturare il database e le prime modalità esplorative era l'obiettivo di avvio di questo progetto. La creazione del database ha posto le basi per nuove riflessioni. Una linea promettente riguarda la possibilità di 'sedimentare' nuove informazioni a partire da quelle già presenti. Un esempio possono essere percorsi tematici, o approfondimenti biografici, in cui è possibile fare riferimento a luoghi, progetti, singoli disegni o persone.

5.5. Conclusioni

L'utilizzo di metodologie provenienti dalla *information visualization* sta conoscendo una diffusione sempre più ampia nella progettazione di interfacce per archivi e collezioni digitali. Abilitare un approccio esplorativo alle informazioni contenute in un archivio eterogeneo richiede un'attenta progettazione fin dalla strutturazione del database, integrando competenze di design della comunicazione. Le interfacce esplorative, grazie alla loro capacità di "mostrare tutto" non solo rendono più veloci e semplici le operazioni di ricerca, ma abilitano azioni prima impossibili, generando nuova conoscenza. Utilizzare i metadati degli elementi per costruirne il contesto offre nuove prospettive su di essi, abilitando un processo di scoperta per un pubblico generico ed uno di approfondimento per studiosi ed esperti del dominio. A fronte delle grandi potenzialità di questo tipo di interfacce, ben espresse dagli studi di Hinton e Whitelaw, un accesso puntuale e analitico resta necessario per gli esperti. Le interfacce esplorative non possono e non devono sostituire completamente le funzioni offerte dal tradizionale campo di ricerca.

L'utilizzo della *information visualization* per offrire accesso al contenuto di collezioni digitali è finora stato utilizzato come strumento per costruire singole viste sull'archivio. Le interfacce sperimentali presentate mostrano come la *information visualization* possa assumere il ruolo di linguaggio visuale coerente in grado di guidare l'utente tra molteplici viste sull'archivio, generando nuova conoscenza sia dalle singole viste, sia dal legame che si crea tra esse tramite l'interazione.

5.6. Bibliografia

- CIUCCARELLI P., INNOCENTI P. (2006). *Sistemi conoscitivi per il design. Una proposta metodologica. Il caso DesignNet*. POLIdesign.
- CORTESE, C. (2011) *Semantic Search and Browsing nell'ambito dei Beni Culturali*. In *Bollettino del CILEA*, no. 118, Consorzio Inter-universitario Lombardo per l'Elaborazione Automatica, pp. 26-30.
- FRIENDLY M. E DENIS D. J. (2006) *Milestones in the history of thematic cartography, statistical graphics, and information visualization*. URL= <http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf> (1 feb. 2013).
- HINTON S., E WHITELAW M. (2010). *Exploring the digital commons: an approach to the visualisation of large heritage datasets*. In *Electronic visualisation and the arts (EVA 2010)*.
- MANOVICH L. (2011) *What is Visualization?* In *Visual Studies*. Routledge.
- MANOVICH L. (2012) *Museum Without Walls, Art History Without Names: Visualization Methods for Humanities and Media Studies*. In *Oxford Handbook of Sound and Image in Digital Media*.
- MARCHIONINI G. (2006) *Exploratory Search: From Finding to Understanding*. In *Communications of the ACM* Vol. 49 no. 4, pp. 41–46.
- RUECKER S. et al. (2012) *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing*. Ashgate Publishing, Ltd.
- SHNEIDERMAN B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*.
- WHITELAW, M. (2009) *Visualising the Digital Archive: the Visible Archive project*. *Archives & Manuscripts*, vol. 37 no. 2, pp. 22–40.
- WHITELAW, M. (2012) *Towards Generous Interfaces for Archival Collections*. In *International Council on Archives Congress*.

6. Convergere a valle. Lo studio del punto di vista degli utenti degli ambienti culturali digitali e l'esperienza del progetto "Una Città per gli Archivi"

*Pierluigi Feliciati*¹

Abstract. Un fruttuoso dialogo tra gli specialisti del trattamento digitale di fonti storiche, archivi, biblioteche e beni artistici, deve includere le problematiche legate alla efficace realizzazione di quegli ambienti ove rendere disponibili per un pubblico più o meno largo le risorse digitali. Questo ambito si scontra di norma con alcune differenze di approccio delle diverse comunità – riguardanti la sostanza, quando non solo la forma – in quanto a logiche descrittive, aggregative, di restituzione, di standard applicati e di profili di utenza previsti. E' interessante, in questo senso, allargare la visuale dalle sole attività a monte dei progetti – in cui le questioni sono modelli, opzioni tecnologiche e ambienti di *output* – al comportamento dei destinatari dei servizi, esaminando quale sia l'effettiva esperienza delle risorse digitali, nell'ottica della condivisione di metodi di indagine e in vista della definizione di metriche di valutazione. In Italia, ma non solo, sono rare le ricerche sistematiche sugli utenti digitali culturali, quindi ogni esperienza di un qualche rilievo metodologico e statistico acquista particolare valore. Partendo dalle questioni metodologiche degli *user studies* per ambienti culturali digitali, si presentano qui i tratti principali dell'articolato studio di valutazione e messa a punto del prototipo del portale "Una Città per gli Archivi" basato sul coinvolgimento degli utenti.

Parole chiave. archivi on line, biblioteche digitali, musei virtuali, user studies, usabilità, esperienza utenti digitali.

¹ Università di Macerata, Dipartimento di scienze della formazione, dei beni culturali e del turismo, Macerata (Italy). e-mail: pierluigi.feliciati@unimc.it.

6.1. Il ruolo degli utenti digitali

Nel 2006, Clay Shirky, una delle voci più interessanti nella comunità professionale dei *network promoters*, propose cinque regole chiave a vantaggio dell'agenzia newyorchese specializzata in *new media* dove operava, "to sort the wheat from the chaff" (Shirky 2006)². I semplici principi sono: 1. *Don't Believe the Hype*, ovvero sono pochissime le innovazioni tecnologiche che hanno davvero cambiato i paradigmi della società, quindi si deve essere sospettosi rispetto all'illusione dell'innovazione continua; 2. *Trust the Would I use it? Test*, ovvero gli utenti, come noi d'altra parte, scelgono i servizi che davvero servono e desiderano fare più comodamente e velocemente le stesse cose che hanno sempre fatto; 3. *Don't Confuse Their Ideas With Your Ideas*, ovvero meglio evitare di basarsi su asserzioni troppo presuntuose, che potrebbero convincere a credere che la nostra esperienza equivale a quella di chiunque; 4. *Information Wants to be Cheap*, ovvero è opportuno non sopravvalutare i costi di acquisto dell'informazione, confondendoli con quelli dei supporti dell'informazione; 5. *It's the Economy, Stupid*, definendo l'economia come l'insieme dei milioni di utenti della Rete, i loro bisogni e le loro effettive disponibilità di tempo e pazienza, tanto che la Rete stessa non può più essere concepita come spazio unicamente tecnologico ma come comunità di persone collegate. Insomma, ormai più di 10 anni fa, rivolgendosi a chi intendesse accingersi a sviluppare prodotti per la Rete, Shirky invitava a preoccuparsi delle reali esigenze degli utenti piuttosto che affidarsi al *common sense* e all'idea poco corretta che la qualità (e il successo) delle nuove tecnologie risieda soprattutto nella creazione di (cari) bisogni innovativi.

Emerge una nuova attenzione alla nozione di utenza dei servizi digitali, non necessariamente schiacciata sulla classica figura del *prosumer* e non solo muovendosi nel *range* delimitato da un lato dall'esaltazione del web spazio di democrazia, frutto del contributo di tutti, dall'altro dallo sfruttamento commerciale sfacciato di questo (abbondante) contributo.

La centralità degli utenti nella progettazione dei sistemi è insistentemente ribadita dagli studi più recenti di *information science*,

² L'autore definisce questa sua lista anche Clay Shirky's Five Rules For Figuring Out When Networking Marketers Are Blowing Smoke.

ritornando in fondo ai sani, vecchi principi del *software engineering*³:

Durante la progettazione e/o la gestione di un servizio di informazione dobbiamo sempre tenere presenti gli utenti, in modo che ogni bit di informazioni memorizzate all'interno del sistema sia accessibile da parte degli utenti interessati, e come corollario gli utenti del sistema informativo o del servizio debbano trovare le informazioni di cui hanno bisogno in modo efficiente ed efficace (Chowdhury, 2011, p. 2-3, traduzione mia).

Venendo al mondo della cultura digitale, comunque declinata, prevale in genere una concezione del valore dei progetti basata unicamente sui contenuti: valore per definizione alto, trattandosi di cultura, buona per definizione, rafforzata dalla quantità e dalla correttezza delle scelte tecniche. La qualità, dunque, nel giudizio corrente (che condiziona spesso l'eventuale supporto finanziario) risiede nel tasso di innovatività, nella ortodossia nella scelta ed applicazione degli standard formali di settore, mentre non sembra influente che si dedichi gran cura all'esito finale degli ambienti digitali culturali, ovvero all'*uso* che ne verrà fatto.

Vale la pena ricordare, ad esempio, quanto notato da (Khoo et al. 2009): "In the case of digital library researchers, the focus of research is often on technical issues (e.g., information retrieval methods, software architecture, etc.) rather than on user-centered issues". In ambito archivistico, poi, Anneli Sundqvist (Sundqvist, 2007, p. 624) aveva già osservato che la "general knowledge of user behaviour is a mixture of common sense, presumptions and prejudice". L'*Institute of Museum and Library Services* aveva da parte sua già osservato, a proposito di coinvolgimento degli utenti nella progettazione di servizi, che "the most frequently-used needs assessment methods do not directly involve the users" (IMLS, 2003, p. 2).

In Italia, non è agevole smentirlo, quando si progettano servizi web culturali, nelle accademie, nei musei, negli archivi e nelle biblioteche, nei centri di documentazione e nelle fondazioni, non ci si basa

³ Principi che affermano, ad esempio, l'importanza capitale dell'analisi dei requisiti di tutti gli *stakeholders*, da effettuarsi ben prima di individuare le possibili soluzioni tecnologiche, oppure che evidenziano l'utilità del *testing* accurato del prototipo, per evitare che una percentuale troppo elevata dei costi sia dedicata alla correzione e manutenzione del prodotto-software già in esercizio. Cfr., tra i tanti manuali, Cohn 2005.

praticamente mai su studi sistematici delle aspettative, dei comportamenti e della soddisfazione finale degli utenti, né si adottano metodi avanzati di simulazione dei profili d'uso. Ammesso che si prendano in considerazione gli utenti, si presume di conoscerli, accontentandoci di proiezioni di noi stessi condite da molto *common sense* e certi dell'entusiasmo che susciteranno i nostri contenuti *well-formed*⁴. Una certa prassi, specie delle pubbliche amministrazioni, nella gestione dei progetti informatici riguardanti la cultura e i beni culturali, esclude sistematicamente che il collaudo dei servizi possa essere condotto coinvolgendo anche gli utenti di quei servizi.

Si fa più profondo, in questo clima, il solco che divide le prassi degli sviluppatori dei sistemi e il mondo della ricerca, con i primi che si ispirano frequentemente a soluzioni standardizzate oppure a sperimentarne di nuove, dando per certo che i comportamenti dei cosiddetti nativi digitali siano tecnologicamente avanzati ed omogenei, e i secondi che raramente coinvolgono gli utenti in modo sistematico e in relazione a questioni specifiche, come la facilità d'uso e l'intuitività dei comandi, l'identificazione dei possibili futuri bisogni, i profili di utilizzo, le somiglianze e le differenze di comportamento tra gruppi di età, origini, status sociale e paesi diversi, l'efficacia nella risposta alle aspettative, l'affidabilità. Inoltre, sarebbe più che mai opportuno non

⁴ Cito ad ulteriore prova di quanto sostenuto (con il consenso degli autori, che ringrazio) un esemplare, recente scambio di mail tra i colleghi Maurizio Lana e Federico Meschini sulla mailing list dell'Associazione di Informatica Umanistica e Cultura Digitale <aiucd-l.humnet.unipi.it>. Il primo osservava, il 20 febbraio 2013, che "il meccanismo dei finanziamenti alla ricerca - lo so per esperienza diretta a livello italiano e europeo - spinge a costruire progetti "grandi" perché questa è sentita come la principale via per mostrare che il proprio progetto è meritevole di attenzione e di finanziamento tanto più se il finanziatore stesso chiede - come sempre - progetti innovativi, progetti avanzati, progetti complessi (...); dall'altra l'ampiezza dei progetti deve cercare di non sconfinare nella complicazione, nella costruzione di un oggetto di interesse che ha la sua sola ragion d'essere nel definire la condizione di esistenza del progetto stesso". Il secondo, il 27 dello stesso mese, aggiungeva come "Da un lato va forse sottolineata la distinzione di ciò che è complesso da ciò che è complicato, intendendo quest'ultimo in senso negativo, mentre il primo come un qualcosa di decisamente non banale e dalle funzionalità avanzate, ma che dall'esterno risulta estremamente semplice da usare. Dall'altro va considerato invece che ciò che per molti è un mezzo per altri è un fine. Per noi in particolare che ci occupiamo di "sistemi informativi digitali orientati verso le scienze umane" (etichetta forse impropria ma che è anche volutamente provocatoria), il fine spesso è il sistema stesso, e qui penso a tutte quelle discipline in cui la forma è predominante".

dimenticare mai la necessaria cura degli aspetti legati alla effettiva accessibilità degli ambienti digitali.

Ha destato molto interesse, qualche anno fa, uno studio britannico (Rowlands et al., 2008) basato sull'analisi avanzata dei *log files*, commissionato dalla British Library e dal JISC allo scopo di sondare il rapporto fra gli utenti e le risorse informative di ambito scientifico in rete. La prima questione affrontata da questo studio è se i cosiddetti *nativi digitali* stiano o meno aggiornando i propri metodi di individuazione dei contenuti, e quanto eventualmente siano diversi da quelli dei ricercatori più anziani. Sono state esaminate prima di tutto le frontiere della *Google generation*, i cui comportamenti sembrano caratterizzati da una tendenza alla ricerca orizzontale dell'informazione, con una forte prevalenza del tempo dedicato alla navigazione rispetto alla lettura, tendenzialmente molto breve. Merita una citazione l'emergere del comportamento "scoiattolante" (*squirreling*) degli utenti, che vede prevalere l'accumulo di materiali scaricabili individuati come interessanti, solo tramite il titolo o l'abstract, nella tana rappresentata dal proprio PC, piuttosto che la loro consultazione nei sistemi. Non è poi possibile verificare se i materiali scaricati vengano usati (ovvero consultati per intero) *offline* (Prensky 2001). Uno studio più recente sugli utenti e le funzionalità del prototipo del portale Europea, attuato in diversi paesi europei e adottando metodologie diversificate, ha aggiunto un ulteriore contributo sia alle risultanze dello studio di Rowlands et al. che a propagandare la sostenibilità degli studi sugli utenti per i progetti, anche non continentali, di *digital library* (Dobrevà et al. 2010a, 2010b; Dobrevà, Feliciati 2010).

In ogni caso, per avviare una più efficace convergenza tra le tante anime della *digital culture*, si può prima di tutto concordare su una comune nozione di *digital library*, come collezione digitale opportunamente gestita e accessibile di risorse, digital-born oppure digitalizzate, testuali, audiovisive, etc. Non va ignorata la complessità di tale nozione, che può sottolineare l'aspetto delle comunità coinvolte, definendo "a (potentially virtual) organization that comprehensively collects, manages, and preserves for the long term rich digital content and offers to its user communities specialized functionality on that content, of measurable quality, and according to prescribed policies" (Candela et al., 2008), oppure ponendo l'accento sull'interazione, definendo le biblioteche digitali come *conversazioni* (AIB, 2005). Certo, sarebbe poco corretto negare che l'*information behaviour* degli utenti è fortemente

dipendente dalle specifiche caratteristiche in termini di origine, forma e logiche di aggregazione delle risorse e di modalità di identificazione, selezione ed uso delle collezioni digitali.

Nonostante ciò, adottare una ampiezza di sguardo più olistica delle problematiche progettuali, condividendo modelli e buone pratiche piuttosto che approfondendo le differenze, non può che giovare all'efficacia finale degli ambienti digitali. Basterebbe ad esempio partire dalla constatazione che una *digital library* è un prodotto software, così da includerla nella definizione fornita dalla norma (ISO 2001), secondo la quale la qualità risiede nella "capacità del prodotto software di consentire agli utenti di raggiungere specifici obiettivi con efficacia, efficienza, sicurezza e soddisfazione in determinati contesti d'uso. Questa definizione sottolinea che la qualità di un prodotto software non sta nell'assenza di errori, nella ricchezza di funzioni, o nel grado di innovazione tecnologica, ma nella misura di cui prodotti sono utilizzabili e accessibili secondo le esigenze degli utenti nei propri contesti di utilizzo.

Muovendo dalla rinnovata attenzione agli studi sugli utenti delle biblioteche digitali, intesa come chiave su cui impernare non solo il successo immediato in termini di *audience* dei sistemi/servizi digitali ma anche per sostenerne lo sviluppo, si espongono qui alcune delle problematiche relative ai modelli e alle problematiche nei diversi contesti di applicazione degli *user studies*, espondendo in conclusione le scelte adottate in una recente ricerca in Italia per la finalizzazione di un portale, svoltasi durante la sua fase prototipale e centrata sul coinvolgimento a vario titolo di campioni di utenza.

6.2. Modelli, contesti e utenti: l'importanza degli *user studies*

Le aree di ricerca sugli utenti sono tradizionalmente occupate dagli studi sull'*information behaviour* (Wilson, 1999) e sull'*user experience* (Hassenzahl, Tractinsky, 2006; Mahlke y Thüring, 2007). Ambedue, pur non occupandosi specificamente di servizi digitali d'informazione, costituiscono un punto di partenza inevitabile per avvicinarsi alle problematiche e alle metodologie adottabili negli studi sugli utenti delle *digital libraries*. Queste linee di ricerca, tuttavia, non offrono un riferimento definitivo e condiviso e tipicamente sono caratterizzate soprattutto dal riconoscimento del valore della ricerca empirica.

Inoltre, si deve guardare anche oltre gli approcci nei confronti degli utenti tipici del marketing (che considerano gli utenti come gruppi omogenei) (Accoto 2007), come l'*Audience metrics*, che nel contesto del web di concentra su indicatori qualitativi e quantitativi per l'analisi di efficacia delle applicazioni. L'*audience* espressa in numeri produce *ratings*: sebbene gli studi sui media tradizionali siano ormai standardizzati (per esempio, i minuti medi di visione di un canale TV) il *new medium* non è più caratterizzabile con il semplice modello di "esposizione", in quanto spazio "interattivo" delle azioni, in cui l'utente non è confinato solo a guardare, ma è invitato a interagire. Un altro recente metodo sviluppato nel contesto degli studi di marketing è il cosiddetto *neuromarketing* (Shiv et al. 2005), che adatta tecniche di analisi neurologiche per indagare comportamenti tipici di utenti/consumatori. Si combinano allo svolgimento di *task* analisi diagnostiche come l'elettroencefalogramma, la risonanza magnetica funzionale, l'*eye-tracking* e l'analisi sistematica della comunicazione non verbale a cura di psicologi cognitivi.

Un contributo interessante viene anche dal mondo del *software engineering*, che ci costringe tra le altre cose a pensare agli utenti di un servizio online non limitandoci ad immaginare solo persone (Felicati, Natale 2008, p. 105-118): lo sviluppo impetuoso del ruolo degli *user agent* (*robot, spider, crawler, harvester...*), ovvero di applicazioni che usano, recuperandoli, memorizzandoli e/o riusandoli, i contenuti e servizi messi a disposizione nella Grande Rete ci costringe a considerare con serietà anche questa tipologia di utenza, valutandone la "soddisfazione" (Nicholas, Clark 2012).

Tra le variabili in gioco nella valutazione delle ricerche sugli utenti, va citata anche la tempistica rispetto al ciclo di vita dei progetti: in che fase è più utile studiare il punto di vista degli utenti, e che tipo di dati è più opportuno raccogliere? Alcune fasi, ad esempio quella di concezione iniziale o quella di valutazione del prototipo, risultano infatti particolarmente critiche perché più opportune per attivare procedure di valutazione delle aspettative; altre, come quella di revisione dell'applicazione dopo un certo tempo, risultano più adatte per verificare la eventuale soddisfazione degli utenti (Felicati, Natale 2008, p. 103-104).

Sembra interessante evidenziare come – pur con alcune mancanze ancora da colmare – i modelli esistenti più spendibili sono quelli elaborati nel campo delle *digital libraries*, tendenzialmente orientati all'usabi-

lità o alla comprensione dell'utenza. Ad esempio, il DELOS *Digital Library reference Model* (Candela et al. 2008) individua l'utente come uno dei sei domini nell'universo delle digital library: *Content, Functionality, User, Architecture, Policy and Quality*. Ogni dominio è decomposto poi in un più fine livello di granularità, definendo per ciascuno il *central concept*, che per il dominio *User* è quello di *Actor*. L'utente finale è uno dei possibili ruoli dell'actor, insieme al DL designer, al DL System Administrator, al DL Application Developer, ma è ulteriormente raffinabile nelle tre classi di *Content creator, Content consumer e Librarian*.

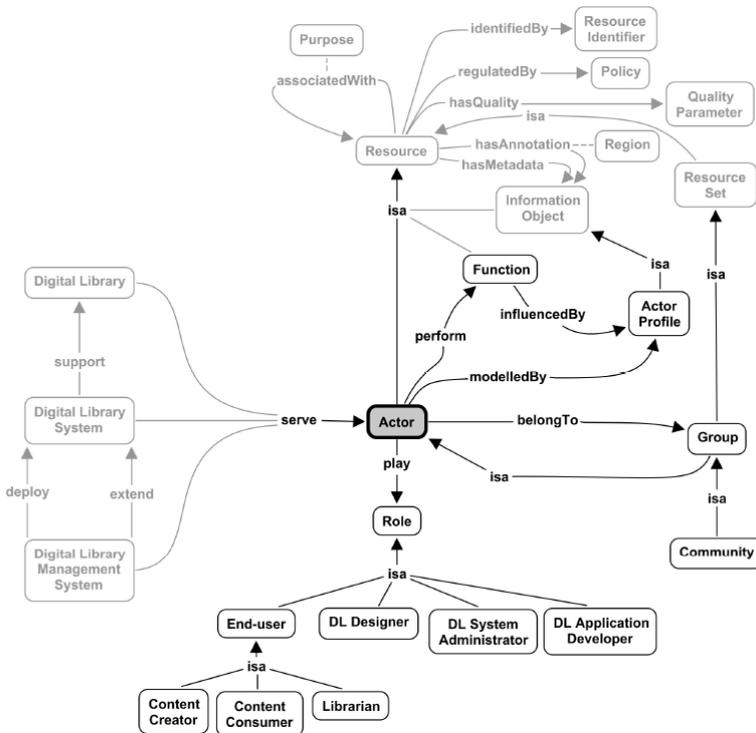


Fig. 6.1. Gli actor in DELOS DLRM.

specificano quella particolare comunità di utenti. Gli sviluppi successivi del modello (Gonçalves et al. 2007) includono la valutazione della digital library, considerando un certo numero di dimensioni qualitative: accessibilità, accuratezza, completezza, componibilità, conformità, coerenza, efficacia, efficienza, estensibilità, pertinenza, conservabilità, pertinenza, affidabilità, riusabilità, significato, somiglianza, e tempestività. Le caratteristiche misurabili sono invece il tempo di risposta (collegato alla efficienza), il costo della migrazione (collegato alla conservabilità) e il numero di disservizi (collegato alla affidabilità). Questo modello, pur non essendo centrato sugli utenti, suggerisce un utile supporto nella definizione di dimensioni qualitative e metriche misurabili estendibili anche all'esperienza degli utenti, ad esempio il tempo di risposta, da collegare non solo all'efficienza del sistema ma anche alle scelte finali degli utenti.

Infine, gli autori del modello 5M (Darányi et al. 2010) sostengono che la attuale complessità delle biblioteche digitali e il loro ampio uso in contesti internazionali obbligano ad affrontare una serie di questioni multiculturali e multilinguistiche. Così, i concetti che caratterizzano le dimensioni attuali per le digital library diventano multiculturalità, multilinguismo e multimodalità. Se questi concetti mostrano in quali ambienti esistono le biblioteche digitali, va aggiunta la sfida di costruire e attuare modelli di biblioteche digitali in grado di affrontare la natura federata delle risorse e le loro dimensioni, in termini di volume di storage necessario ma anche in termini di numero di oggetti che compongono le collezioni digitali. Il modello 5M propone insomma di affinare il modello per i contenuti a partire da tre specifiche di contenuto (multimodalità, multilinguismo, multiculturalità), costruendo un ponte tra DELOS DLRM e 5S, con una granularità più fine delle caratteristiche delle biblioteche digitali e tenendo in considerazione non solo gli oggetti digitali, ma anche i contesti. Si propone insomma di:

"aggiungere un livello 5M altamente integrato con il modello 5S, ad un livello inferiore, dove 5M sta per collezioni digitali Multiculturali, Multilingue, con contenuti Multimodali, indicizzati, suddivisi in categorie e recuperate da algoritmi Multivariati, sfruttando la loro semantica in continua evoluzione attraverso la Modellazione delle collezioni e degli utenti" (traduzione mia).

L'attuazione pratica di questo modello continuerà ad evolversi, ma sembra riassumere alcune delle sfide più importanti del settore: mettere a fuoco quali sono davvero i contenuti delle biblioteche digitali, in modo che i modelli possono accoglierne al meglio le complessità e valorizzare la rappresentazione dei contesti.

Due particolari accezioni di ambienti digitali culturali, particolarmente di successo negli ultimi anni, meritano qualche notazione aggiuntiva. Prima di tutto, gli ambienti di aggregazione di contenuti provenienti da fonti diverse, i portali, ad esempio Europeana e la World Digital Library. Riunendo diversi tipi di contenuti predisposti in condizioni diverse si aprono nuovi problemi per la fornitura di servizi affidabili in condizioni di omogeneità, di per sé apparentemente impossibili da risolvere. D'altro canto, indipendentemente dalle dimensioni e dal profilo di un biblioteca digitale, è sempre d'aiuto capire meglio gli utenti e le loro esigenze. Anche nel caso di biblioteche digitali caratterizzate da contenuti specifici, in cui le comunità-utente potrebbero sembrare chiaramente definite con le loro esigenze e preferenze, la natura stessa del mondo digitale comporta utenti inaspettati e usi sorprendenti (Miller 2001):

"Per gli utenti, un portale è sicuramente utile solo se effettivamente incontra una necessità di un qualche utente, e in un modo per loro confortevole. In quanto tale, il portale deve offrire di più rispetto a qualsiasi degli altri servizi disponibili. Per garantire ciò, è necessario un lavoro costante per garantire l'interoperabilità dei sistemi" (traduzione mia).

Un'altro ambito particolare da considerare è quello della realtà aumentata e dei modelli virtuali, dove pure gli utenti e i loro effettivi comportamenti non sono considerati con particolare attenzione. Come bene riassumono (Champion, Sekiguchi 2004): "there has so far been little research into evaluation best suited for evaluating and improving the experience and learning of participation in a virtual environment and even less work done on virtual heritage environment."

I modelli virtuali di rappresentazione del patrimonio culturale non sono creati per essere utilizzati isolatamente, ma per avere davvero il valore di un'esperienza culturale devono essere contestualizzati e collegati con altri oggetti. Un meccanismo naturale per raggiungere

questo obiettivo è associarli ai contenuti delle biblioteche digitali. Da questo punto di vista, è importante capire non solo ciò di cui gli utenti avrebbero bisogno e cosa si aspettino da un modello virtuale, ma come queste applicazioni possono essere efficacemente integrate con altri contenuti culturali.

Per concludere, un ambito particolarmente delicato in quanto a restituzione dei contenuti in ambienti digitali è senza dubbio quello archivistico, caratterizzato da sempre per il peso del ruolo di mediazione specialistica svolto dall'archivista, pur con la disponibilità di strumenti di ricerca completi e dettagliati, basati sull'applicazione rigorosa degli standard internazionali (Felicati, Alfier 2012). Emergono con evidenza, dai pochi studi e ricerche empiriche sulla qualità dell'esperienza degli utenti rispetto agli archivi online (Chapman, 2009 e 2010; Daniels, Yakel, 2010; Duff, 2012; Duff, Stoyanova, 1998; Scheir, 2005; Yakel, 2004), alcune problematiche tipiche che potremmo suddividere in linguistiche, logico-sintattiche e di sistema (Felicati, 2007). Per quanto riguarda le prime, si nota che le interfacce di navigazione e le descrizioni archivistiche sono spesso comunicate con un linguaggio specialistico, auto-referenziale. Raffinati distinguo storico-amministrativi e tecnico-istituzionali, di cui non sempre è facile individuare le chiavi d'interpretazione, condizionano fortemente la logica degli ambienti archivistici digitali: mancando la soggettazione, le titolazioni delle entità descrittive (le originarie denominazioni delle antiche magistrature che hanno prodotto gli archivi e delle loro ripartizioni interne) sono infatti l'unica chiave per la navigazione delle strutture ad "albero rovesciato" e la fonte indicizzata per le ricerche testuali.

Le criticità logico-sintattiche riguardano la restituzione agli utenti delle strutture dei fondi: è inevitabile orientarsi entro costruzioni gerarchiche talvolta molto complesse, difficili da sciogliere talvolta dagli stessi archivisti, elaborate secondo criteri imposti da contesti "naturalmente" particolari, perché dovuti all'attività specifica dei corrispondenti soggetti produttori. In sostanza, si sceglie di sottolineare la sintassi del sistema di comunicazione a svantaggio della chiarezza e della reperibilità dei contenuti. Infine, sono significative le criticità di sistema, quelle derivanti dall'ampiezza del dominio informativo che si rende disponibile per l'utente: il *coverage*, ovvero "the spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is rele-

vant"(DCMI Metadata terms, 2012). L'utente infatti non è sempre messo nelle condizioni di focalizzare con chiarezza alcune premesse fondamentali alla sua esperienza di fruizione: valutare se il sistema archivistico prescelto sia il più idoneo per la sua ricerca, intuendo se ciò che lo interessa possa essere o meno descritto con esso; comprendere nel caso se l'oggetto d'interesse non possa essere recuperato in alcune risorsa esistente della rete; infine, giudicare se la propria ricerca sia o meno pertinente.

Questa sintetica classificazione trova un preciso riscontro sperimentale in alcune ricerche sugli utenti svolte in contesto nord-americano – già citate più sopra – in cui oltre alle definizioni astratte si è voluto *ask the users*, ovvero prevedere attività sistematiche di raccolta del feedback degli utenti, facendo tesoro della valutazione dei dati per eventuali interventi sui sistemi. Si è costruito, e condiviso, un piccolo patrimonio di studi sugli utenti degli archivi online basato sull'applicazione di differenti metodi di analisi. Tali ricerche convergono, comunque, nell'individuazione di alcune tipiche criticità d'uso: la terminologia archivistica, tendenzialmente troppo tecnica⁵; la restituzione della struttura gerarchica delle descrizioni archivistiche, su cui si focalizza troppo l'interesse degli operatori d'archivio quando invece gli utenti pongono in primo piano soprattutto l'*aboutness* degli archivi (Duff, Johnson, 2002, p. 477); le funzioni di ricerca, il cui uso consapevole e produttivo richiede spesso competenze troppo elevate; la qualità delle descrizioni, infine, che per alcuni utenti risultano troppo prolisse, per altri troppo sintetiche.

Per concludere, se da molte parti viene avvertita la necessità di migliorare la qualità del processo progettuale degli ambienti digitali culturali, lo stato dell'arte appare ancora povero di esperienze, disomogeneo e poco coordinato. Di certo, l'adozione di modalità aperte di pubblicazione dei risultati e delle metodologie potrebbe favorire la creazione di network e, in prospettiva, sostenere la formalizzazione condivisa di metriche e di schemi metodologici, così da poter finalmente confrontare diverse ricerche.

⁵ Duff y Stoyanova, 1998, p. 59-60 affermano ad esempio che "it is sufficient quoting the term 'fonds', the core concept of archival profession, which sounds often ambiguous or even incomprehensible to users".

6.3. La ricerca sul prototipo di *Una Città per gli Archivi*

Un'esperienza che può costituire un buon riferimento applicativo, in questo quadro, è rappresentata dal progetto "Una città per gli archivi", varato nel 2007 per iniziativa della Fondazione del Monte di Bologna e Ravenna e della Fondazione Cassa di risparmio di Bologna con lo scopo di contribuire alla conservazione e accessibilità dei fondi legati alla storia bolognese, prodotti perlopiù tra il XIX e il XX secolo⁶. Il complesso progetto è stato svolto con il supporto scientifico di storici e archivisti di chiara fama e con il coinvolgimento attivo – non solo di facciata – di diversi uffici e strutture cui è affidata la tutela e valorizzazione degli archivi in area bolognese: la Soprintendenza archivistica per l'Emilia-Romagna, l'Archivio di Stato di Bologna, l'Istituto per i beni ambientali culturali naturali dell'Emilia-Romagna (IBC), la Fondazione Gramsci per l'Emilia-Romagna, la Provincia di Bologna e la Biblioteca comunale dell'Archiginnasio.

Gli archivi oggetto degli interventi di tutela, restauro, inventariazione e riproduzione digitale sono quanto mai vari in quanto a soggetti produttori, contesti storici e tipologie dei supporti, per un totale di più di 200 fondi, con l'individuazione e descrizione di circa 300.000 unità descrittive e la digitalizzazione di 200.000 documenti: basta ricordare i fondi connessi con le stragi degli anni Ottanta del Novecento, prodotti e raccolti dalle associazioni delle vittime del 2 agosto e di Ustica⁷ o le serie sui sovversivi della Questura di Bologna⁸, o gli archivi di sindacati, enti assistenziali e scuole, o ancora le carte personali di Marco Minghetti, fino a nuclei speciali come quello dei film amatoriali e di famiglia, digitalizzati e conservati presso l'Associazione *Home Movies*.

Il progetto sta avviandosi ormai alla sua conclusione con la pubblicazione in rete, prevista per la primavera del 2013, di un portale web dedicato alla restituzione pubblica dell'ingente materiale informativo

⁶ Sul progetto cfr. soprattutto Alfier 2012 e Antonelli 2012.

⁷ Cfr. su questi archivi il convegno *Archivi negati archivi "supplenti"*. Le fonti per la storia delle stragi e del terrorismo, Bologna, giugno 2011, www.sa-ero.archivi.beniculturali.it/index.php?id=783.

⁸ La serie dell'Archivio di Stato di Bologna, *Questura di Bologna, Gabinetto*, categoria A8 "Persone pericolose per la sicurezza dello Stato", 1872-1983, è stata oggetto di una giornata di studio, "Male qualità". Controllo di polizia e azione giudiziaria tra Otto e Novecento nelle carte dell'Archivio di Stato di Bologna", Bologna, ottobre 2010.

prodotto. Nelle intenzioni dei coordinatori del progetto questo portale dovrà costituire un medium semplice e intuitivo per navigare nella storia recente di Bologna, uno strumento di facile comprensione e di immediata consultazione, pensato per un pubblico che si vorrebbe eterogeneo per competenze, interessi ed esigenze di ricerca. La scelta, infatti,

"è stata quella di progettare un portale dinamico, che a partire dalla consueta consultazione della struttura ad albero, multilivellare e gerarchica, dei complessi documentari inventariati (*browsing*), offrisse anche la fruizione delle copie digitali relative a fotografie, manifesti, materiale iconografico, filmati e fonti sonore, e mettesse a disposizione degli utenti meno esperti di ricerca su inventari d'archivio modalità d'interrogazione degli inventari basate su liste di parole chiave (come nomi di persona, di luogo, di autore e come voci di soggettazione, per tema o argomento). La dinamicità del portale permetterà di seguire percorsi ontologici in grado di evidenziare relazioni inaspettate per l'utente, di proporre mostre virtuali ideate dalla redazione o proposte e realizzate da enti o singoli ricercatori, di pubblicare contributi di ricerca di natura scientifica o didattica, di esaminare i documenti di progetto, di avere a disposizione un ambiente digitale che solleciterà i "naviganti" a salvare le proprie ricerche in un'area loro riservata dove potranno integrarle, ri-organizzarle e, qualora lo ritengano opportuno, condividerle con altri" (Antonelli 2012, p. 8-9).

Un prototipo del portale sufficientemente stabile, almeno con riferimento alle funzioni centrali di ricerca e di presentazione dei risultati, navigazione degli alberi gerarchici e visualizzazione delle descrizioni era già disponibile nel settembre del 2011. Su proposta di chi scrive e con il coordinamento scientifico condiviso con Alessandro Alfier, coordinatore tecnico del progetto, si è stabilito di affiancare al normale collaudo a cura dei coordinatori tecnici e scientifici e dell'azienda responsabile dello sviluppo un progetto per il suo assestamento e valutazione basato, appunto, sul coinvolgimento degli utenti. Il primo passo, alla base di questa tutto sommato nuova ricerca sul suolo italico, è stata l'attività comune svolta nell'ambito del master sulla formazione, gestione, conservazione degli archivi digitali in

ambito pubblico e privato dell'Università di Macerata⁹, occasione ideale per attivare una collaborazione fruttuosa ed impostare lo staff di progetto.

La articolata ed ampia ricerca basata sugli utenti dell'efficacia di dettaglio del prototipo del portale *Una Città per gli Archivi*, che qui si presenta solo dal punto di vista delle premesse, degli obiettivi e della metodologia, è già stata oggetto di alcuni interventi in convegni internazionali¹⁰, mentre i suoi risultati finali (ovvero i dati raccolti, le elaborazioni statistiche e le suggestioni metodologiche da queste derivate) saranno presto resi pubblici in rete, così da poter fungere anche da ispirazione per successive ricerche simili nel nostro paese¹¹.

Una premessa importante da sottolineare concerne la fase del progetto nella quale si è intervenuti con la ricerca basata sugli utenti, quella di assestamento del prototipo: ovviamente non è stato possibile investigare sulle reali esperienze degli utenti finali nell'uso dei contenuti e delle funzionalità del sistema. Nondimeno, in questa delicata fase precedente il rilascio pubblico, in cui committenti e sviluppatori verificano il prodotto finale e lo sottopongono agli ultimi aggiustamenti, gli studi sugli utenti sono preziosi per raccogliere dati che sappiano guidare gli interventi di assestamento e sviluppo anche di *release* successive.

⁹ Questo Master che vede coinvolte diversi Dipartimenti dell'UNiversità di Macerata (Scienze della Formazione Beni culturali e Turismo, Economia e Diritto e Studi Umanistici), diretto da Stefano Pigliapoco, è giunto nell'a.a. 2012/13 alla sua VI edizione. Il *project work* di Alessandro Alfier, svoltosi nell'edizione 2010/11 e sfociato in una interessante relazione, è stata supervisionato da chi scrive, in qualità di docente dell'insegnamento "Dati e metadati, dagli standard archivistici ai profili applicativi gestionali".

¹⁰ Ovvero l'intervento di chi scrive e di Alessandro Alfier, *Archives on the Web and users expectations: towards a convergence with digital libraries* alla 7th SEEDI Conference 2012, Digitisation of cultural and scientific heritage, tenutasi a Lubljana (17-18 maggio 2012), http://www.nuk.uni-lj.si/dokumenti/2012/seedi/5_3_Presentazione_Seedi-AlfierFeliciatidef.pdf e quello di chi scrive, *Testing of an archival DL prototype asking to users: the case of "Una città per gli Archivi"* al convegno ACM/IEEE Joint Conference on Digital Libraries - JCDL 2012, nell'ambito del tutorial su *User studies for digital library development*, a cura di Milena Dobрева, Pierluigi Feliciati e Nicci Westbrook (Washington, 10-14 giugno 2012), <http://www.slideshare.net/pfeliciati/testing-a-a-dl-prototype>.

¹¹ Sulla pubblicazione in rete di Open Scientific Data, cfr. tra gli altri Cribb, Sari 2010.

Le domande che sorgono più frequentemente, a proposito degli *user studies*, sono: quali metodologie adottare, per ottenere quali dati e coinvolgendo quali e quanti utenti? E quali e quante risorse è stato necessario investire?¹² La strategia di ricerca adottata, possibile solo grazie all'entusiastico e valido supporto organizzativo e tecnico delle due Fondazioni, è stata principalmente quella di attivare un pacchetto differenziato di metodologie di studio, cercando al tempo stesso di investigare diversi campioni di utenza e puntando così a raccogliere un'importante quantità di dati sia qualitativi che quantitativi. Dal punto di vista delle risorse, potendo contare su un'ottima rete di contatti già costituiti e su un valido management, si è verificato che non sono stati necessari grandi investimenti, anche percentualmente rispetto al budget complessivo del progetto¹³ e nonostante la ricerca sia stata tanto articolata diacronicamente e logisticamente.

Prima di tutto, si è scelto di partire dall'utenza interna al progetto, raramente coinvolta se non "artigianalmente", nella valutazione dell'ambiente di output. Si sono infatti consultati, organizzando un apposito *focus group*, diverse classi di utenti interni, mettendo insieme 20 persone con *background* professionali e ruoli nel progetto molto differenziati: il management di progetto, un campione degli archivisti responsabili della produzione dei contenuti inventariali destinati alla pubblicazione nella *digital library*, tutti gli archivisti già coinvolti nell'analisi delle funzioni della piattaforma ed infine gli informatici responsabili dello sviluppo del portale.

Per questo studio iniziale, il cui incontro si è svolto nel novembre 2011, è stato elaborato un protocollo specifico che riunisse le metodologie tipiche del *focus group* per il *testing* di prodotti con quelle del brainstorming di progetto. E' stata prevista la moderazione di chi scrive, soggetto esterno al percorso pluriennale di "Una città per gli archivi", dunque indipendente rispetto alle scelte effettuate, con l'obiettivo di orientare l'indagine su alcuni focus: le particolari aspet-

¹² Cfr. Dobрева, Feliciati, O'Dwyer 2012, part 5, p. 247-253.

¹³ Pur non potendo effettuare una valutazione esatta, credo comunque di poter approssimare che le ricerche sugli utenti del prototipo abbia pesato meno dello 0,05% sul budget complessivo dell'intero progetto.

tative degli autori degli inventari archivistici, dunque di contenuti specialistici, rispetto alla restituzione sul web, messe a confronto con le interfacce-utente degli sviluppatori, verificando in tal modo se tutte le potenzialità fossero state esplorate e quali possibili criticità fossero ancora risolvibili. Il protocollo adottato per questo *focus group* è stato articolato in quattro sezioni, per un totale di 24 domande e relative discussioni:

1. questioni generali relative alla *home page*;
2. questioni relative alle funzioni di navigazione della struttura ad albero degli inventari;
3. questioni inerenti alle funzioni di ricerca e alle liste di risultati;
4. questioni attinenti alla struttura delle descrizioni archivistiche.

Vista la ricchezza ed utilità dei risultati di questo *panel group* interno, si è ritenuto utile metterne in calendario uno equivalente in prossimità del rilascio del portale, allo scopo di verificare la qualità dei lavori svolti in un anno di *assessment* e l'eventuale cambiamento di prospettive critiche da parte degli archivisti coinvolti nel progetto e dei tecnici autori dell'ambiente di restituzione.

Il *core* del progetto di ricerca con il supporto degli utenti, comunque, è stato costituito dalla serie di cinque *focus group*, in cui sono stati riuniti campioni significativi di possibili utenti finali del portale in via di rilascio. Tutti i gruppi di discussione moderata sono stati basati sullo stesso dettagliato protocollo¹⁴, che prevedeva diversi questionari, sezioni di discussione guidata e lo svolgimento di alcuni specifici compiti da compiere usando il prototipo di portale. Nello specifico, le sezioni del protocollo adottato sono:

- un questionario demografico e di confronto tra le competenze informatiche dichiarate con quelle effettive;
- una breve presentazione degli scopi e delle funzionalità del prototipo di portale;
- la prima raccolta di impressioni sul portale, basata su una discussione generale;

¹⁴ Questo protocollo è parzialmente ispirato a quello adottato dal gruppo di ricerca sul prototipo di Europeana 1.0, *User and functionalities testing*, su cui cfr. Dobрева et al. 2010.

- l'esecuzione di cinque *task* che obbligassero all'uso effettivo del prototipo, i cui risultati sono stati registrati attraverso la compilazione di specifici questionari composti da diverse domande per ciascuna delle attività;
- una raccolta attraverso discussione e domande di opinioni avanzate, proposte e critiche fondate sull'uso del prototipo di portale;
- una discussione finale guidata sull'utilità del portale e possibili sviluppi del progetto.

Ogni sessione di *focus group* è stata coordinata da due moderatori (uno esterno ed uno interno al progetto, con ruoli diversificati e definiti nel dettaglio,) ed è durata circa 100 minuti. Più di 60 persone in totale hanno partecipato alle sessioni di analisi, di cui 15 studenti delle scuole superiori (età 16-19 anni), 16 studenti universitari provenienti da diversi corsi di laurea, divisi in due diverse sessioni (una a Macerata e una a Bologna, età 23-29 anni), 16 persone classificabili come *general public* (età 39-63 anni, professori di scuola superiore, impiegati con mansioni amministrative, un architetto, un web master, tre pensionati) ed infine 16 professionisti nel campo dei beni culturali (età 29-52 anni: archivisti, bibliotecari ed esperti di storia del cinema). Le sessioni, documentate anche con riprese fotografiche grazie al consenso esplicito dei partecipanti, sono state tenute sempre in aule informatiche connesse alla Internet, a Macerata nel polo informatico di palazzo Ciccolini e a Bologna presso il polo tecnico-scientifico degli Istituti Aldini-Valeriani.

Il terzo polo della ricerca, infine, previsto ma ancora non attuato, è costituito da un *expert study* (Klas 2012), finalizzato ad una valutazione specializzata e di dettaglio delle funzionalità del portale e basato sul coinvolgimento di sette esperti cui sarà sottoposto a distanza un questionario comprensivo di valutazioni d'insieme e seguenti allo svolgimento di specifici *task*. Questa metodologia di studio è stata rinviata alla fase immediatamente precedente il rilascio definitivo dell'applicazione, forse addirittura a portale già pubblicato, perché è indispensabile che le considerazioni degli esperti siano svolte esaminando un ambiente avanzato, sia in termini di completezza e validazione dei dati che rispetto all'attivazione di tutte le funzionalità. Tra gli esperti contattati, tre saranno archivisti non coinvolti direttamente nel progetto (di cui due professori universitari e un

direttore di Archivio di Stato), tre esperti di *digital library* e di metadati in ambito culturale ed infine un professore universitario di sistemi di elaborazione delle informazioni specializzato in *human-computer interaction*. Il protocollo adottato per questa linea di studio sarà focalizzato sulle performance, l'efficacia delle funzionalità e delle soluzioni adottate, sull'usabilità in termini stretti, sulla terminologia e sulla qualità dei contenuti, ispirandoci al modello fornito dal *Tryptich Interaction Framework* (Tsakonas, Papatheodorou 2008; Tsakonas 2012).

Si prevede di concludere l'elaborazione finale dei dati raccolti nella primavera del 2013, soprattutto intersecando le risultanze delle diverse metodologie adottate così da valorizzarne al massimo gli esiti. L'analisi dei dati, già utilizzati dai coordinatori del progetto durante tutto il processo di *finalization* del prototipo di portale, servirà a comporre un quadro delle principali criticità che, a partire dalla piattaforma di "Una città per gli archivi", presentano gli archivi on line sotto l'aspetto della qualità dell'accesso e dell'uso. Da un lato il management del progetto avrà una base informativa dettagliata per programmare gli interventi di manutenzione evolutiva del portale, già in esercizio, dall'altro si spera di offrire utili spunti di riflessione sui comportamenti degli utenti, ad uso della comunità nazionale degli archivisti e dei gestori di biblioteche digitali.

6.4. Bibliografia

- ACCOTO C. (2007). *Misurare le audience in internet. Teorie, tecniche e metriche per la misurazione degli utenti in rete*, Franco Angeli.
- ALFIER A. *La descrizione archivistica in "Una città per gli archivi": metodi e strumenti*. «Archivi & Computer», n. 2/2012, pp. 37-68.
- ANTONELLI A. *Istantanea "Una città per gli archivi": istituzioni, fatti, persone, tempi, modi, prospettive e storia di un progetto archivistico locale*. «Archivi & Computer», n. 2/2012, pp. 7-36.
- ASSOCIAZIONE ITALIANA BIBLIOTECHE - GRUPPO DI STUDIO SULLE BIBLIOTECHE DIGITALI. *Manifesto per le biblioteche digitali, 2005-12*, URL= <http://www.aib.it/aib/cg/gbdigd05a.htm3> [ultima visita 11.3.2013].

- CANDELA L., CASTELLI D., FERRO N., IOANNIDIS Y., KOUTRIKA G., MEGHINI C., PAGANO P., ROSS S., SOERTEL D., AGOSTI M., DOBREVA M., KATIFORI A., SCHULDH H. (2008). *The DELOS Digital Library reference model - Foundations for digital libraries*. Version 0.98. 2008, URL= http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf [ultima visita 11.3.2013].
- CHAPMAN J. C. (2009). *What would users do? An empirical analysis of user interaction with online finding aids*, Master's paper submitted to the faculty of the School of information and library science of the University of North Carolina. Chapel Hill.
- CHAPMAN J.C. (2010). *Observing users: an empirical analysis of user interaction with online finding aids*. «Journal of archival organization», n. 8, p. 4-30.
- CHAMPION E., SEKIGUCHI S. (2004). *Cultural Learning in Virtual Environments*. In VSMM2004 Hybrid Realities: Digital Partners, Ogaki City, Gifu, Japan, 17-19 November 2004, pp. 364-373, URL= <http://www.itee.uq.edu.au/~erikc/papers/> [ultima visita 11.3.2013].
- CHOWDHURY G., CHOWDHURY S. (2011). *Information users and usability in the digital age*. Facet Publishing.
- COHN M. (2005). *Agile Estimating and Planning*, Prentice Hall.
- CRIBB J.,SARI T. (2010). *Open Science: Sharing Knowledge in the Global Century*, Csiro Publishing.
- DANIELS M. G., YAKEL E. (2010). *Seek and you may find: successful search in online finding aid systems*. «American archivist», n. 73, p. 535-568.
- DARÁNYI S., WITTEK P., DOBREVA M. (2010) *Position paper: adding a 5M Layer to the 5S Model of Digital Libraries*. In Digital libraries for international development, Surfers Paradise, in conjunction with Joint JCDL/ICADL.
- DUBLIN CORE METADATA INITIATIVE (2012), *DCMI Metadata Terms. Coverage*, URL= <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=elements#coverage> [ultima visita 11.3.2013].
- DOBREVA M., MCCULLOCH E., BIRRELL D., FELICIATI P., RUTHVEN I., SYKES J., ÜNAL Y. (2010). *User and Functional Testing. Final report*, URL= <http://version1.europeana.eu/web/europeana-project/documents> [ultima visita 11.3.2013].

- DOBREVA M., MCCULLOCH E., BIRRELL D., FELICIATI P., RUTHVEN I., SYKES J., ÜNAL Y. (2010). *Digital Natives and Specialised Digital Libraries: a Study of Europeana Users*. In S. Kurbanoglu et al. (a c. di), *Technological Convergence and Social Networks in Information Management*. Springer-Verlag, p. 45-60.
- DOBREVA M., FELICIATI P. (2011). *User-centric evaluation of cultural heritage digital libraries: three case studies*. «NCD Review. Pregled nacionalnog centra za digitalizaciju», n. 18, p. 1-14, URL= http://elib.mi.sanu.ac.rs/pages/browse_issue.php?db=ncd&rbr=18 [ultima visita 11.3.2013].
- DOBREVA M., O'DWYER A., FELICIATI P. (a cura di) (2012). *Users Studies for Digital Library Development*, Facet Publishing.
- DUFF W. (2012). *User studies in Archives*. In Dobreva, O'Dwyer, Feliciati (a cura di), *User studies for digital libraries development*, Facet Publishing, p. 199-206.
- DUFF W., DRYDEN J., LIMKILDE C., CHERRY J., BOGOMAZOVA E. (2008). *Archivists' Views of User-based Evaluation: Benefits, Barriers, and Requirements*. «American Archivist», 71/1, p. 144-166.
- DUFF W., JOHNSON C. A. (2002). *Accidentally found on purpose: information-seeking behavior of historians in archives*. «Library quarterly», n. 72 (4), p. 472-496.
- DUFF W., STOYANOVA P. (1998). *Transforming the crazy quilt: archival displays from users's point of view*. «Archivaria», n. 45, p. 44-79.
- FELICIATI P., *Dall'inventario alla descrizione degli archivi in ambiente digitale: si possono offrire agli utenti risorse efficaci?* (2007). In "Elaborare il sapere nell'era digitale", Montevarchi (Italy), 22-23 novembre 2007, URL= <http://eprints.rclis.org/11159/> [ultima visita 11.3.2013].
- FELICIATI P., NATALE M.T. (a cura di) (2009). *Manuale per l'interazione con gli utenti del web culturale*, ed. it., MINERVA EC Working Group 5, URL= http://www.minervaeurope.org/publications/handbook_webusers_it.htm [ultima visita 11.3.2013].
- FELICIATI P. (2009). *I requisiti di fattibilità di un sistema informativo archivistico: modelli organizzativi, informatici e soddisfazione degli utenti*. «Archivi», a. IV, n. 1 e URL= <http://eprints.rclis.org/12834/> [ultima visita 11.3.2013].

- FELICIATI P., ALFIER. (2012) *Archives on the web and users expectations: towards a convergence with digital libraries*. Paper for the 7th SEEDI Conference on digitisation, access and preservation of cultural and scientific heritage in the South-Eastern European countries, Ljubljana, Slovenia 17-18 May 2012. « NCD Review. Pregled Nacionalnog Centra za Digitalizaciju», vol. 22, c.s.
- GONÇALVES M.A., FOX E.A., WATSON L. T., KIPP N.A. (2004). *Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries*. «ACM Transactions on Information Systems», 22(2), p. 270-312.
- GONÇALVES M.A., MOREIRA B.L., FOX E.A., WATSON L.T. (2007). *What is a good digital library? - A quality model for digital libraries*. «Information Processing & Management», vol. 43, fasc. 5, p. 1416-1437.
- MARC HASSENZAHN, NOAM TRACTINSKY. *User experience – a research agenda*. Behaviour & information technology n. 25 (2), 2006, p. 91-97.
- INSTITUTE OF MUSEUM AND LIBRARY SERVICES (2003). *Assessment of End-User Needs in IMLS-Funded Digitization Projects*, 41 pp. URL= <http://www.imls.gov/assets/1/AssetManager/userneedsassessment.pdf> [ultima visita 11.3.2013].
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2010). *ISO 9241. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*.
- KHOO M., BUCHANAN G., CUNNINGHAM S.J. (2009). *Lightweight user-friendly evaluation knowledge for digital libraries*. «D-Lib Magazine», Vol. 15 N. 7/8, URL= <http://www.dlib.org/dlib/july09/khoo/07khoo.html> [ultima visita 11.3.2013].
- KLAS C-P.(2012). *Expert evaluation methods*. In M. Dobрева, A. O'Dwyer, P. Feliciati (a c. di), *User studies for digital libraries development*, Facet Publishing, p. 75-84.
- MAHLKE S., THÜRING M. (2007). *Studying antecedents of emotional experiences in interactive contexts*. In Proceedings of computer /human interaction 2007 conference, ACM Press, p. 915-918.
- MILLER P. (2001). *The concept of the portal*, «Ariadne», n. 30, URL= <http://www.ariadne.ac.uk/issue30/portal/> [ultima visita 11.3.2013].

- NICHOLAS D., CLARK D. (2012). *Evidence of user behaviour: deep log analysis*. In M. Dobрева, A. O'Dwyer, P. Feliciati (a c. di), *User studies for digital libraries development*, Facet Publishing, p. 85-94.
- PRENSKY M. (2001). *Digital natives, digital immigrants*. «On the Horizon», 9/5, pp. 1-6, URL= <http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20P%20art1.pdf> [ultima visita 11.3.2013].
- ROWLANDS I., NICHOLAS D., WILLIAMS P., HUNTINGTON P., FIELDHOUSE M., GUNTER B., WITHEY R., JAMALI H.R., DOBROWOLSKI T., TENOPIR C. (2008). *The Google generation: the information behaviour of the researcher of the future*. «Aslib Proceedings», vol. 60 Issue 4, pp. 290 - 310.
- SCHEIR W. (2005). *First entry: report on a qualitative exploratory study of novice user experience with online finding aids*. «Journal of archival organization», n. 3 (4), p. 49-85.
- SHIRKY C. (2006). *Don't believe the hype*. In Clay Shirky's Writings About the Internet Economics and Culture, Media and Community, Open Source, URL= <http://www.shirky.com> [ultima visita 11.3.2013].
- SHIV B., BECHARA A., LEVIN I., ALBA, BETTMAN J.R., DUBE J.R., ISEN A., MELLERS B., SMIDTS A., GRANT S.J. (2005). *AP Decision Neuroscience*. «Marketing Letters», 16 (3/4), p. 375-386.
- SUNDQVIST A. (2007). *The use of records: a literature review*. «Archives and Social studies: a Journal of Interdisciplinarity Research», 1 (1), p. 623-653.
- TSAKONAS G., PAPTAEODOROU C. (2008). *Exploring usefulness and usability in the evaluation of open access digital libraries*. «Information Processing & Management», vol. 44, fasc. 3, p. 1234-1250.
- TSAKONAS G. (2012). *Users within the evaluation of digital libraries*. In M. Dobрева, A. O'Dwyer, P. Feliciati (a c. di), *User studies for digital libraries development*, Facet Publishing, p. 51-61.
- WILSON T.D. (1999). *Models in information behaviour research*. «Journal of documentation» n. 55 (3), p. 249-270.
- YAKL E. (2004). *Encoded archival description: are finding aids boundary spanners or barriers for users?* «Journal of archival organization», n. 2 (1/2), p. 63-77.

7. Una risorsa online per la storia dell'arte: il database della Fondazione Federico Zeri

*Francesca Mambelli*¹

Abstract. Il progetto con cui la Fondazione Federico Zeri ha avviato la trasposizione in digitale dell'archivio fotografico donato dallo studioso all'Università di Bologna ha permesso di realizzare un catalogo informatizzato di immagini e metadati oggi considerato uno dei più importanti database al mondo sull'arte italiana, consultabile gratuitamente online sul sito della Fondazione. L'articolo descrive brevemente l'origine e la composizione del fondo, analizza le scelte metodologiche e gli standard utilizzati per la sua catalogazione e digitalizzazione, riassume le caratteristiche principali della banca dati; si sofferma infine sulle prospettive future del progetto che vanno nella direzione dell'integrazione tra record catalografici diversi (fotografici, storico-artistici, bibliografici e archivistici) e di una maggiore interazione tra utenti e sistema.

Parole chiave: fotografia, catalogazione, digitalizzazione, interattività, archivio integrato.

¹ Fondazione Federico Zeri/ Università di Bologna, Bologna, Italia.
e-mail: francesca.mambelli6@unibo.it.

7.1. La raccolta analogica: censimento e filologia

Il rapporto che lega Federico Zeri e la fotografia è ben esemplificato da questa citazione tratta dall'autobiografia *Confesso che ho sbagliato*, in cui lo studioso espone il metodo di lavoro dello storico dell'arte-conoscitore, identificandone con precisione gli 'strumenti del mestiere':

Comincio col guardare ripetutamente le innumerevoli fotografie che mi vengono recapitate, esaminandole dapprima nell'insieme e poi, con una lente, nei dettagli. Preciso che le fotografie debbono essere in bianco e nero: anche se può sembrare un paradosso, non riesco a leggere correttamente le fotografie a colori dove ogni dato è affogato in una sorta di minestrone; le riproduzioni a colori impediscono di isolare le forme, di analizzare lo stato di conservazione della superficie, che è la prima cosa che faccio con le fotografie in bianco e nero. In seguito viene l'identificazione del soggetto, che è parte integrante del quadro, indissolubile dalla forma (...). Proseguendo cerco di stabilire la regione, la data e la mano dell'artista, basandomi sui dati stilistici (Zeri 1995, 125-126).

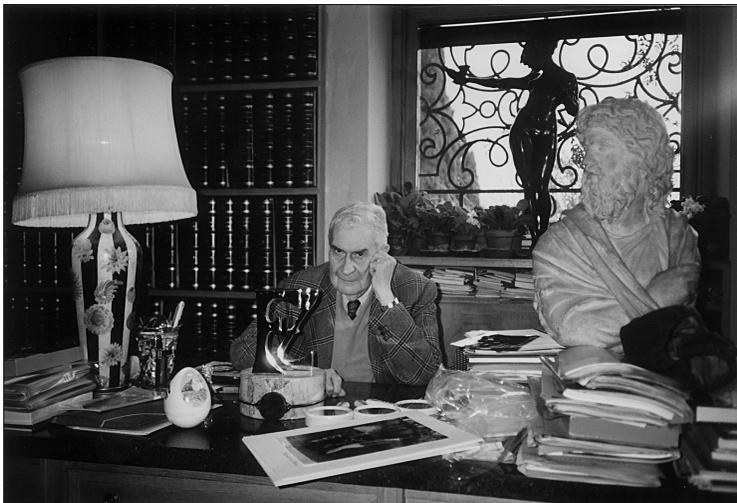


Fig. 7.1. Federico Zeri nel suo studio a Mentana, tra i libri e le fotografie.

Federico Zeri, come gli altri grandi storici del secolo scorso, fondeva la sua analisi filologica sulla consultazione quotidiana delle fotografie che venivano spostate continuamente sul tavolo di lavoro, accostando dettagli, soggetti, ricomponendo sequenze, per venire poi archiviate in una raccolta organica. Un metodo di indagine che confidava sulla sensibilità dell'occhio, continuamente allenato sui documenti visivi, nella convinzione di Bernard Berenson che nella storia dell'arte "vince chi ha più fotografie". Un metodo che, fino alla rivoluzione costituita dall'avvento del digitale, ha prodotto archivi fotografici quali quelli di Zeri, Longhi, Ragghianti, Briganti, Berenson (solo per citarne alcuni), capaci di racchiudere in sé sia la storia di una disciplina – cioè la storia dell'arte – sia gli sforzi tecnologici con cui la fotografia ha cercato di illustrarla e diffonderne la conoscenza².

Strumento di lavoro quotidiano per Federico Zeri era dunque la fototeca: una raccolta di 290.000 fotografie da lui organizzata nel corso degli anni come una struttura complessa e diramata in grado di trasmettere non solo il pensiero dello studioso sulle singole opere ma la sua idea stessa di storia dell'arte. Iniziato alla fine degli anni Quaranta e accresciuto con passione nell'arco di tutta la vita, l'archivio fotografico di Zeri spazia dall'antichità all'età moderna, dalla pittura italiana e straniera alla scultura e al disegno, dalla natura morta alle arti decorative. Le fotografie, circa il 40% delle quali databili entro gli anni Trenta del Novecento, documentano opere oggi disperse o distrutte, in localizzazioni originarie o in stati conservativi precedenti a restauri storici.

I fototipi hanno la provenienza più disparata: dalla foto singola inviata da un collezionista, ai materiali arrivati in grandi stock grazie ai contatti con case d'asta, Soprintendenze e Gabinetto Fotografico Nazionale; dagli archivi di fotografi e antiquari a fondi appartenuti a storici dell'arte fra cui Antonio Muñoz, Umberto Gnoli, Guglielmo Matthiae e Evelyn Sandberg Vavalà (Sambo, Culatti 2011, 19).

Nelle mani di Zeri queste raccolte perdevano l'originaria fisionomia e venivano smembrate attraverso un capillare sistema di archiviazione. Il materiale è infatti distribuito in circa 1.000 buste che recano l'intestazione del secolo e dell'area geografica (Secolo XIV: Umbria; Secolo XV: Veneto; etc.), secondo un'impostazione che riprende gli stu-

² Sul rapporto tra fotografia e storia dell'arte, in particolare sul ruolo degli archivi fotografici negli studi storico-artistici, si vedano i contributi raccolti nel fondamentale volume C. Caraffa, a c. di (2011) e bibliografia citata.

di di un pioniere della storia dell'arte come Luigi Lanzi (Cavicchi 2011, 30); all'interno delle scatole sono contenute cartelle intestate alle diverse individualità artistiche, in un ordine cronologico e di importanza che parte dalle personalità di maggiore spicco per proseguire con l'attività della bottega, gli autori minori, gli anonimi e le opere problematiche.

Le fotografie recano sul retro indicazioni autografe di Zeri e dei precedenti possessori che trasmettono importanti informazioni circa la storia attribuzionistica o collezionistica dell'opera e offrono riferimenti incrociati con la biblioteca. La stessa architettura tassonomica della fototeca, per secoli, aree geografiche, personalità e scuole, e la presenza di note autografe e appunti caratterizzano anche la biblioteca dello studioso: 47.000 volumi di storia dell'arte raccolti parallelamente alle foto, con le quali instaurano continui rimandi. A questi due importanti nuclei della donazione di Zeri - biblioteca e archivio fotografico - si affianca la collezione dei cataloghi d'asta, che, con i suoi 37.000 esemplari, rappresenta la più importante raccolta presente in Italia dedicata a questa tipologia di materiali.



Fig. 7.2. L'archivio fotografico nei locali climatizzati del ex Convento di S. Cristina, Bologna.

7.2. Il progetto di catalogazione e digitalizzazione

Alla morte dello studioso, avvenuta il 5 ottobre 1998, l'Università di Bologna, erede delle sue collezioni, costituisce la Fondazione Zeri con l'obiettivo prioritario di valorizzare questo lascito e di tutelare, divulgare, promuovere l'opera e la figura di Federico Zeri. Dopo il trasferimento dei materiali a Bologna si procede quindi alla catalogazione dei volumi in SOL e all'apertura al pubblico della biblioteca nei locali dell'ex convento di Santa Cristina.

Nel 2003 prende avvio il progetto sull'archivio fotografico. Un progetto che prevede la digitalizzazione e catalogazione delle fotografie e il cui fine ultimo è, accanto alla messa in sicurezza e alla conservazione degli originali, la costituzione di un database accessibile online che renda disponibile agli studiosi questo ricchissimo materiale.

L'avvio della catalogazione è stato preceduto da una fase di riflessione scientifica ed elaborazione teorica che ha preso in considerazione sia l'aspetto degli standard sia il problema degli utenti a cui lo strumento creato si sarebbe indirizzato. Occorre infatti adottare una scheda che fosse compatibile con gli standard catalografici nazionali per i beni fotografici ma anche strutturata in base alla specificità e all'ordinamento del fondo Zeri. L'obiettivo era quello di realizzare un catalogo informatizzato di fotografie (il bene realmente e fisicamente posseduto) consultato prevalentemente da storici dell'arte, il cui approccio sarebbe avvenuto nella maggior parte dei casi a partire dai dati delle opere documentate. La necessità di salvare e normalizzare i molteplici livelli di informazioni presenti sul verso della foto, unita a quella di rendere percepibile la struttura gerarchica dell'archivio Zeri a chi effettuasse una consultazione, rendeva la scheda F ICCD un modello descrittivo insufficiente.

E' stato dunque creato uno standard composto da due unità catalografiche primarie, la scheda *Opera d'arte* e la scheda *Fotografia*. Collegate tra loro in un rapporto 1 a n, esse danno vita ad un database simmetrico e speculare che permette una ricerca bifronte da parte degli utenti e una navigazione trasversale (Mambelli 2011, 45).

La scheda *Opera d'arte* si riferisce al contenuto semantico del documento. E' composta da 80 campi (semplici e strutturati) che consentono di registrare nel dettaglio e rendere ricercabili informazioni quali tipo di oggetto, soggetto, attribuzioni, cronologia, localizzazioni attuali e precedenti, materia e tecnica, bibliografia di riferimento, mostre, documentazione allegata. Per la compilazione di queste voci

la fonte primaria è costituita dalle annotazioni di Zeri e dai dati deducibili dall'ordinamento. La catalogazione è intesa quindi come operazione filologica che dà vita non a un database 'neutro' sull'arte italiana ma che trasferisce dal piano cartaceo a quello digitale un archivio di persona con caratteristiche di grande coerenza interna.

La scheda *Fotografia* prevede 113 campi divisi in paragrafi che si riferiscono all'autore fotografo, all'ubicazione della fotografia, al luogo e data di scatto, alla cronologia del positivo, al soggetto della foto, ai dati tecnici ecc. Si tratta del livello inventariale della scheda F arricchito da due paragrafi del livello precatalogo: il paragrafo *Conservazione* che, compilato e aggiornato, permette di monitorare lo stato conservativo dell'archivio; il paragrafo *Altre localizzazioni* che consente di ricostruire a livello virtuale gli archivi di provenienza delle fotografie, smembrati da Federico Zeri all'interno della propria raccolta.

Giacomo di Nicola da Recanati, San Flaviano				LOCALIZZAZIONE GEOGRAFICO-AMMINISTRATIVA	
CODICI				Stato Italia	
Numero scheda 19652				Regione Toscana	
Numero busta 0233				Provincia Firenze	
Intestazione busta Pittura italiana sec. XV Marche. Gotico; Pietro di Domenico, Giacomo di Nicola, Cristoforo di Giovanni, anonimi				Città Firenze	
Numero fascicolo 2				Contenitore Moretti	
Intestazione fascicolo Giacomo di Nicola, Bartolomeo Frignisico				LOCALIZZAZIONI PRECEDENTI	
OGGETTO				Stato Stati Uniti d'America	
Definizione oggetto scomparto di polittico				Stato federale New York	
Identificazione elemento d'insieme				Città New York (NY)	
Tipologia dipinto				Contenitore Asta Sotheby's	
Soggetto/ Titolo San Flaviano				Data di uscita 30 maggio 1991, n. 3	
Materia e tecnica tavola				Stato Stati Uniti d'America	
Tipo misure cm				Stato federale New York	
Altezza 93				Città New York (NY)	
Larghezza 31				Contenitore Asta American Art Association	
AUTORE				Specifiche vendita Chiesa, 22 novembre 1927, n. 92	
Nome Giacomo di Nicola da Recanati				Stato Italia	
Ambito culturale Scuola italiana, scuola marchigiana				Regione Lombardia	
Motivazione attrib. Attribuzione F. Zeri (14 luglio 1996)				Città Milano	
ALTRE ATTRIBUZIONI				Contenitore Collezione A. Chiesa	
Nome Crivelli Carlo				Data di uscita 1927	
Motivazione attrib. Asta American Art Association (1927)				Stato Italia	
Nome Maestro di Staffolo				Regione Marche	
Motivazione attrib. Asta Sotheby's (1991)				Provincia Macerata	
CRONOLOGIA				Città Recanati	
Secolo sec. XV				Contenitore Cattedrale di S. Flaviano	
Da 1443				BIBLIOGRAFIA	
A 1443				Genere bibliografia specifica	
RIFERIMENTO AD ALTRE SCHEDE				Autore Laciotte M./Mognetti E.	
Schede correlate Scomparti di un polittico smembrato dipinto da Giacomo di Nicola per il duomo di Recanati				Titolo Avignon, Musée du Petit Palais. Peinture italienne	
DOCUMENTI ALLEGATI				Anno di edizione 1967	
Scheda allegata vai alla scheda				Pagine specifiche p. 104	
				Genere bibliografia specifica	
				Autore De Marchi A.G./Angelelli W. (a cura di)	
				Titolo Pittura dal Duecento al primo Cinquecento nelle fotografie di Girolamo Bombelli	
				Anno di edizione 1991	
				Pagine specifiche p. 157	
				FOTO	
				Foto Bombelli G., Giacomo di Nicola da Recanati - sec. XV - San Flaviano - insieme	
				Foto Anonimo, Giacomo di Nicola da Recanati - sec. XV - San Flaviano - insieme	
				Foto Anonimo, Giacomo di Nicola da Recanati - sec. XV - San Flaviano - insieme	

Fig. 7.3. Una scheda opera nel database della Fototeca Zeri.

Bombelli G. Giacomo di Nicola da Recanati, sec. XV San Flaviano		TITOLO Titolo attribuito Giacomo di Nicola da Recanati - sec. XV - San Flaviano Specifiche del titolo del catalogatore
CODICI Tipo scheda F Livello ricerca I Numero cat. generale 51642 Ente schedatore Fondazione Federico Zeri - Università di Bologna	AUTORE FOTOGRAFO Autore fotografo Bombelli, Girolamo Motivazione attrib. timbro Indicazione del nome Studio Fotografico. Specializzato per la riproduzione a colori e in nero di G. Bombelli. Milano - Via Mantegna, 6 - Telef. 60480 - 60481	AUTORE OPERA FOTOGRAFATA Nome Giacomo di Nicola da Recanati Ambito culturale Scuola italiana, scuola marchigiana
LOCALIZZAZIONE GEOGRAFICO-AMMINISTRATIVA Regione 51644 Provincia BO Comune Bologna Contenitore Ex convento di S. Cristina Spazio bibliotec. piazzetta G. Morandi, 2 Raccolta Fondazione Federico Zeri	DATI TECNICI Indicazione di colore BN Materia e tecnica gelatina ai sali d'argento/ carta baritata Misure mm 262x104 Tipo misure supporto primario	CRONOLOGIA Secolo XX Da 1920 ca. A 1940 ca. Motivazione data analisi storica/ analisi tecnico-formale
UBICAZIONE Fondo Fototeca Zeri Serie Pittura italiana Numero busta 0233 Intestazione busta Pittura italiana sec. XV, Marche. Gotico: Pietro di Domenico, Giacomo di Nicola, Cristoforo di Giovanni, anonimi Numero fascicolo 2 Intestazione fascicolo Giacomo di Nicola, Bartolomeo Frignisco Collocazione Pl_0233/2/16 Numero inventario 51642	DATA E LUOGO DELLA RIPRESA Data 1920-1930/ ca.	CONSERVAZIONE Stato conservazione discreto
OGGETTO Definizione postivo Numero oggetti 1 Natura bibliotec. m	RAPPORTO Opera iniziale negativo Collocazione Roma/ Istituto Centrale per il Catalogo e la Documentazione: Fototeca Nazionale N. inventario C 19334	CONDIZIONE GIURIDICA Indicazione generica proprietà Ente pubblico non territoriale Indicazione specifica Alma Mater Studiorum - Università di Bologna
SOGGETTO Soggetto San Flaviano - dipinto	ANNOTAZIONI Osservazioni Fino alla sua ricomparsa sul mercato negli anni Novanta, il dipinto era conosciuto solo grazie a questa fotografia di Bombelli (cfr. De Marchi A/S/ Angelilli W. (a cura di), Pittura dal Duecento al primo Cinquecento nelle fotografie di Girolamo Bombelli, 1991, p. 157).	
COMPILAZIONE Data 2005/12/20 Nome Mambelli F.	SCHEDA OPERA RELATIVA Opera Giacomo di Nicola da Recanati - San Flaviano	

Fig. 7.4. Una scheda F nel database della Fototeca Zeri.

A ciascuna scheda opera sono collegate le schede delle fotografie che la documentano. Ogni scheda foto è illustrata da due immagini digitali, rispettivamente la scansione del recto e del verso del fototipo, con l'alta e la bassa risoluzione collegate rispettivamente alle schede in intranet e in internet. I parametri di digitalizzazione seguono la Normativa per l'acquisizione digitale delle immagini fotografiche dell'ICCD che, per quanto datata, costituisce ancora un buon compromesso tra leggibilità dell'immagine e gestibilità della mole di files prodotti³. Per l'hosting, il back-up e la conservazione a lungo termine di immagini e metadati, la Fondazione si avvale del Centro servizi informatici dell'Università di Bologna.

In fase di progettazione del modulo di data entry e di gestione archivio sono stati parallelamente individuati, per ciascuna voce catalografica, i thesauri e i vocabolari condivisi a livello internazionale

³ <http://www.iccd.beniculturali.it/getFile.php?id=291>.

(in particolare i vocabularies del Getty Research Institute) a cui vincolare la compilazione. Sono stati inoltre realizzati 5 archivi relazionati per alcuni dei nodi informativi che confluiscono nella scheda: authority files dedicati rispettivamente agli autori delle opere d'arte, ai fotografi, alla bibliografia, alle fonti e, infine, alla struttura stessa dell'archivio.

L'esportabilità in xml e l'azione di mapping dei dati sui modelli utilizzati a livello internazionale e nazionale (Dublin Core, EDM, PICO) hanno consentito alla Fondazione di avviare importanti collaborazioni con realtà quali Europea, Artstor e CulturaItalia, nei cui portali è confluita parte delle schede e delle immagini della fototeca Zeri.

L'applicazione di questo modello alle 140.000 fotografie di pittura italiana costituenti il fondo principale dell'archivio ha prodotto un catalogo digitale considerato tra i più importanti database al mondo sull'arte italiana, consultato nel 2012 da circa 65.000 visitatori. Questo risultato si deve, a giudicare dal feedback che riceviamo dai nostri utenti, non tanto alla quantità delle immagini che lo costituiscono, alla rarità del materiale iconografico o ad un'interfaccia utente particolarmente intuitiva, quanto alla capillarità delle informazioni che vengono fornite a corredo delle immagini e alla loro affidabilità scientifica.

Il catalogo digitale, consultabile gratuitamente dal sito della Fondazione⁴, permette interrogazioni immediate e veloci su tutte le opere d'arte e le fotografie presenti nella banca dati. I tre livelli di ricerca, libera, semplice e avanzata, sono possibili sia a partire dai dati della fotografia sia da quelli dell'opera d'arte. Un motore di ricerca interno consente di richiamare ed ordinare i risultati non solo secondo parametri relativi all'opera d'arte ma anche secondo l'autore, il soggetto, la datazione, la tecnica della fotografia. Dall'elenco, dalla gallery o dalla slideshow, è possibile entrare nelle schede e transitare in qualsiasi momento dalla scheda opera a quelle dei relativi fototipi, e viceversa.

Un catalogo così strutturato apre molteplici possibilità di ricerca che interessano non solo la storia dell'arte ma anche discipline ad essa connesse quali la storia della fotografia, la storia della critica e la storia del collezionismo.

⁴ <http://www.fondazionezeri.unibo.it>

7.3. Prospettive future

La catalogazione dell'archivio fotografico di Zeri procede attualmente con il nucleo dedicato alla Natura morta, composto da 14.000 fotografie. Viste le caratteristiche di questo genere pittorico, si è resa necessaria una modifica della scheda catalogafica, con conseguente adattamento del software. In particolare sono state adottate nuove norme per la soggettazione che puntano ad una descrizione iconografica più analitica. In dipinti quasi sempre anonimi e di difficile collocazione spazio-temporale, gli elementi della composizione rappresentano spesso dati discriminanti, in grado di supportare un'attribuzione o di identificare una determinata opera all'interno del contesto. Per questo si è scelto di affiancare al titolo generico l'elenco tutti gli oggetti raffigurati adottando un apposito thesaurus che organizza le parole chiave in categorie, sottocategorie e singoli elementi (es: strumenti musicali; strumenti a fiato; flauto).

Parallelamente la Fondazione ha aperto il progetto a nuovi sviluppi che vanno da un lato verso una maggiore interazione tra la comunità degli utenti e il sistema; dall'altro verso il superamento dei tradizionali confini che separano il trattamento delle varie tipologie di beni documentari.

Per quanto riguarda il primo aspetto, nell'interfaccia del catalogo sul nuovo sito, la cui messa in linea è prevista per metà marzo 2013, saranno inserite funzionalità aggiuntive. Tra queste la possibilità per l'utente di segnalare, per ciascuna scheda, aggiornamenti circa l'attuale collocazione (magari di opere disperse), nuove attribuzioni, bibliografia riguardante l'opera o la fotografia. Informazioni il cui inserimento sarà comunque subordinato a un controllo scientifico da parte degli operatori. Un apposito tasto renderà più agevole la richiesta di riproduzioni, facilitando l'ottenimento di quelle immagini cedibili in quanto prive di limitazioni legate al copyright.

Per quanto riguarda il secondo aspetto, e cioè il dialogo tra sistemi catalogafici diversi all'interno di uno stesso ambiente, è in atto l'estensione della catalogazione ai documenti inseriti da Zeri nelle cartelle della fototeca a corredo della documentazione iconografica. Si tratta di lettere, appunti, estratti da riviste o volumi, bozze di saggi o ancora relazioni di restauro e perizie, connessi fisicamente alle foto e

nel contenuto alle opere d'arte documentate. Per la catalogazione di questo materiale è stato creato un modulo apposito interno alla piattaforma, approcciabile sul web come archivio autonomo. I 21 campi di cui si compone il tracciato soddisfano le norme ISAD(G) e consentono di riportare sia le informazioni sul contesto e sulla posizione gerarchica all'interno dell'archivio, sia una descrizione particolareggiata del contenuto del documento. Ciascuna scheda è corredata dalla scansione, visibile su intranet qualora non contenga dati sensibili.

Un ulteriore progetto riguarda la collezione dei cataloghi d'asta a cui abbiamo precedentemente accennato. La sfida in questo caso riguarda la possibilità di far dialogare il sistema biblioteconomico di SBN e la banca dati dell'archivio fotografico Zeri; in particolare collegando i record bibliografici del singolo catalogo d'asta alle schede di tutte le opere passate in quella determinata vendita e offrendo in questo modo, uno spoglio del volume per lotti.

La modalità individuata per raggiungere l'obiettivo è stata la creazione di un ulteriore modulo interno alla piattaforma della Fondazione in cui vengono importate le notizie dei cataloghi da SBN attraverso i formati di scarico Unimarc. La suddivisione in campi delle aree di descrizione biblioteconomica permette la normalizzazione dei dati, la loro implementazione e il collegamento con schede e immagini dell'archivio fotografico.

Sul database online Cataloghi d'asta sarà possibile effettuare interrogazioni attualmente impossibili tramite SBN: per data in formato standard, luogo di vendita, codice d'asta, elenco delle tipologie di oggetti venduti e delle collezioni. Attraverso il collegamento con le immagini della fototeca sarà possibile 'illustrare' il catalogo, e aggiungere dati e informazioni essenziali per la ricerca storico-artistica, relative ai singoli lotti.

File Modifica Visualizza Preferiti Strumenti ?

FONDAZIONE ZERI | CATALOGO | Allegato : Galleria ...

FONDAZIONE FEDERICO ZERI

Homepage English version

Fondazione Federico Zeri Fototeca Biblioteca Epigrafi

2: Italian primitives and Renaissance paintings: a small group of canvases by flemish and french masters

SCHEDA

N. identificativo	U803820721
Localizzazione	CA 16 1926 0416
Inventario esemplare	5750
Primo autore	*American 'art' galleries
Pubblicazione	New York : American art association
Luogo vendita	New York
Data	1926/04/16
Collezionisti	Achilleo Chiesa
Tipologia oggetti venduti	pitture

SCHEDA COLLEGATE

Opera [Visualizza tutte le schede](#)

Stampa [Ultima ricerca](#) [Nuova ricerca](#)

© 2003 - 2012 Fondazione Federico Zeri - Realizzazione Software: Infrastruttura e Palettforma - Lingua: Italiano pagina

Fig. 7.5. Catalogo della vendita Chiesa del 16/04/1926 nel modulo *Cataloghi d'asta*.

Search Works for

- Maestro della Madonna Lazzarini, *Madonna con Bambino e angeli*
- Pietro di Giovanni, *Crocifissione di Cristo*
- Albertinelli Mariotto, *Madonna con Bambino, san Giovanni, san Francesco...*
- Domenico di Michelino, *Madonna con Bambino*
- Maestro di Santa Cecilia, *Madonna con Bambino in trono e santie - San Mich...*
- Simone di Filippo, *Crocifissione di Cristo - Dio Padre benedicente*
- Andrea di Bartolo, *Madonna con Bambino e angeli*
- Nezzeri Jacopo, *Cristo risorto*
- Maestro di San Pier Damiano, *Visitatione*
- Giacomo di Nicola da Becanati, *San Pleviano*
- Giovanni di Ser Giovanni, *I Nummi Intorno a Totocolo*
- Giovanni del Ricordo, *San Giovanni Battista - Storie della vita di san Gio...*
- Gaddi Arnolfo, *Madonna con Bambino in trono e angeli - Storie della vita di...*
- Paolo Veneziano, *Eremita di san Nicola di Bari*
- Paolo Veneziano, *Nascita di san Nicola di Bari*
- Solaris Andrea, *Salome con la testa di san Giovanni Battista*
- Vivanti Bartolomeo, *Croto in piedi e angeli*
- Lotto Lorenzo, *Ritratto di gentiluomo*
- Anonimo veneto-bizantino sec. XIV, *Deposizione di Cristo dalla croce*
- Jacopo di Paolo, *Madonna con Bambino e angeli*
- Barbanti Lazzaro, *Madonna con Bambino*
- Tranaianni Nicolo', *Giovane pastore*
- Daddi Bernardo, *Madonna con Bambino, san Giovanni Gualberto, san Fanciu...*
- Nicolo' di Seena, *Madonna con Bambino*
- Jacopo di Arancello, *Madonna con Bambino*
- Anonimo padovano sec. XV, *Madonna con Bambino*
- Aspochiani Barnaba, *Incoronazione di Maria VerGINE*
- Cristofano di Papi dell'Altissimo, *Ritratto di Enrico VIII*
- Leonardo da Vinci, *Madonna con Bambino, san Giovanni, un angelo e donat...*
- Tiepolo Giovanni Battista, *Vessovo in prebistera*

page 1 of 2 - results from 1 to 30 of 54

[] options

Fig. 7.6. Opere vendute nell'asta Chiesa del 16/04/1926 presenti nel catalogo online (gallery).

Il sistema che la Fondazione ha inteso creare attraverso la catalogazione delle proprie collezioni è ben esemplificato dal reticolo della figura 7.7 che lo traduce in una struttura 'cellulare'. Al centro del reticolo, e cioè di questa complessa serie di rimandi e di legami, sta la scheda dell'opera, nucleo aggregatore delle informazioni, ma anche unica scheda 'virtuale' di tutto il sistema, cioè relativa al solo oggetto non fisicamente in possesso della Fondazione.

Non un portale indifferenziato in cui convergono, a livello paratattico, materiali e notizie di varia natura, ma un insieme organizzato attorno allo stesso nucleo concettuale ordinatore della mente di Zeri, centro del suo interesse e del suo studio: l'opera d'arte.

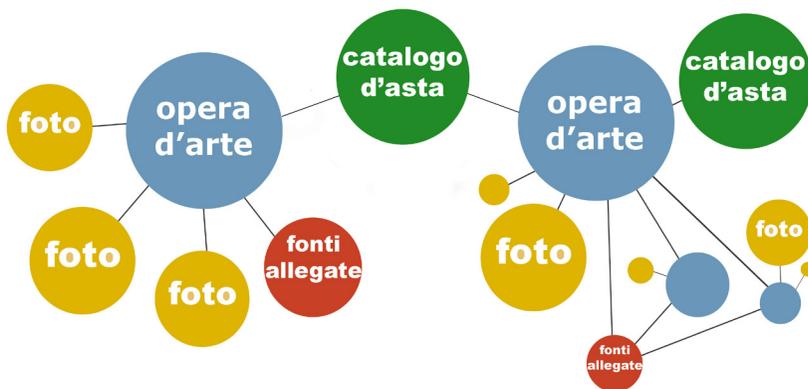


Fig. 7.7. Schema del database online della Fondazione Zeri: unità di accesso e relazioni.

7.4. Bibliografia

- AA.VV. (2008). *Progetto per una fototeca*. «Acta Photographica. Rivista di fotografia, cultura e territorio», a. III, no. 1, Bononia University Press.
- BONETTI M.F. (2001). *La catalogazione dei beni fotografici nella normativa del Ministero per i Beni e le Attività Culturali*. In A. Stanzani, O. Orsi, C. Giudici, a c. di, *Lo spazio, il tempo, le opere. Il catalogo del patrimonio culturale*, Silvana, pp. 188-191.
- CARAFFA C., a c. di (2011). *Photo Archives and the Photographic Memory of Art History*, Deutscher Kunstverlag.

- CAVICCHI M. (2012). *Tra "ecumeniche retate di fotografie" e riordino classificatorio, l'archivio fotografico di Federico Zeri*. «Predella», no. 32. URL=http://predella.arte.unipi.it/index.php?option=com_content&view=article&id=277&catid=86&Itemid=113. [ultima visita 19.2.2013].
- CAVICCHI M. (2011). *Per un censimento della pittura italiana: i cataloghi di Federico Zeri*. In A. Ottani Cavina, a c. di, *La pittura italiana nella Fototeca Zeri. Fotografie scelte: Lazio, Toscana, Marche, Umbria, Emilia Romagna*, Umberto Allemandi & C., pp. 30-36.
- GIUDICI C., a c. di (2004). *C'era due volte. Fondi fotografici e patrimonio artistico*, Minerva.
- GIUSA A., a c. di (2003). *Fotografia e patrimonio culturale: nuove tecnologie per la valorizzazione della storia per immagini*, Comune di Verona.
- MAMBELLI F. (2011). *La Fototeca Zeri: da archivio privato a banca dati online*. In A. Ottani Cavina, a c. di, *La pittura italiana nella Fototeca Zeri. Fotografie scelte: Lazio, Toscana, Marche, Umbria, Emilia Romagna*, Umberto Allemandi & C., pp. 42-46.
- OTTANI CAVINA A., a c. di (2009). *Federico Zeri, dietro l'immagine. Opere d'arte e fotografia*, Umberto Allemandi & C.
- SAMBO E., CULATTI M. (2011). *La pittura italiana nella Fototeca Zeri*. In A. Ottani Cavina, a c. di, *La pittura italiana nella Fototeca Zeri. Fotografie scelte: Lazio, Toscana, Marche, Umbria, Emilia Romagna*, Umberto Allemandi & C., pp. 17-24.
- SPIAZZI A.M., MAJOLI L., GIUDICI C., a c. di (2010). *Gli archivi fotografici delle soprintendenze. Tutela e storia. Territori veneti e limitrofi*, Terra Ferma.
- ZERI F. (1995). *Confesso che ho sbagliato. Ricordi autobiografici*, in collaborazione con P. Mauriès, Longanesi & Co.

PARTE III

FILOLOGIA DIGITALE: ESPERIENZE E SPERIMENTAZIONI

8. Stemma codicum: analisi e generazione semi-automatica

Gioele Barabucci, Angelo Di Iorio, Fabio Vitali¹

Abstract. Questo articolo presenta un metodo semi-automatico per la ricostruzione dello stemma codicum a partire da una collezione di documenti XML. A differenza dei metodi esistenti, il metodo proposto permette allo studioso di filologia di personalizzare e controllare il processo di ricostruzione in maniera dichiarativa, consentendo l'indicazione di quali sono gli errori più rilevanti e la definizione di nuovi tipi di errori. Questo approccio porta ad una virtuosa ripartizione dei compiti: gli esperti di informatica implementano sistemi per gestire le parti meccaniche (confronti tra documenti, ricerca degli errori, calcoli di affinità) lasciando ai filologi la parte creativa del processo (selezione degli errori più significativi, definizione di nuove classi di errori, integrazione con informazioni non testuali).

Parole Chiave. stemma codicum, XML/TEI, filologia digitale, XML diff algorithm.

¹ Università di Bologna. e-mail: barabucci, diiorio, vitali@cs.unibo.it.

8.1. Introduzione

La costruzione di uno stemma codicum è un processo complesso e laborioso che richiede elevate competenze nel campo della filologia e dell'ecdotica. L'esperienza, la conoscenza dei codici, l'intuizione e le capacità di deduzione sono fattori determinanti per costruire uno stemma.

Sostituire l'analisi di un filologo con strumenti completamente automatici è una soluzione impossibile a nostro avviso. D'altra parte alcune attività di questo processo possono essere automatizzate. Si pensi, ad esempio, alla mera individuazione di differenze tra testimoni (diversa dalla loro successiva interpretazione e analisi) o alla verifica di diverse configurazioni dello stemma (spesso simulate attraverso modelli matematici).

Questo articolo presenta un approccio innovativo per generare lo stemma che demanda a una applicazione il calcolo delle differenze e l'analisi sintattica dei testi, e permette allo studioso di focalizzarsi sulla parte filologica, di definire le proprie tecniche in aggiunta a quelle esistenti e di sperimentare varie ipotesi di lavoro, senza sforzi manuali per la riscrittura dello stemma.

Le idee di base sono state applicate con profitto al calcolo delle differenze e degli emendamenti tra documenti legislativi; è altrettanto interessante studiarne applicabilità, potenzialità e limiti nel campo degli studi umanistici. Sebbene la ricerca sia nelle fasi iniziali, i primi risultati sembrano promettenti e aprono interessanti scenari e spunti di riflessione.

8.2. Stemma codicum, edizioni critiche e codifica XML

Il metodo proposto in questo articolo presuppone la codifica in XML dei testimoni. La marcatura di documenti in XML è largamente apprezzata come strumento di rappresentazione, archiviazione e preservazione dei testi, come illustrato, tra gli altri, in DeRose 1999. I vantaggi di un linguaggio di markup descrittivo rispetto ad un testo semplice sono molteplici: possibilità di catturare il ruolo strutturale delle varie parti, possibilità di esprimere diverse strutture su uno

stesso documento (anche se con tecniche complesse), possibilità di separare aspetti strutturali e logici da aspetti puramente presentazionali, possibilità di modificare gli aspetti presentazionali e visualizzare gli stessi contenuti in molteplici formati, possibilità di creare documenti fruibili sia da utenti umani che da componenti software e implementare strumenti sofisticati di analisi, e così via.

Non sono mai state sperimentate, tuttavia, soluzioni per generare stemma a partire da documenti codificati in XML. L'idea di base di questo lavoro, invece, è usare algoritmi di calcolo automatico delle differenze tra documenti XML per individuare gli errori presenti tra i vari testimoni, analizzare questi errori e classificarli, e sfruttare questa analisi per generare lo stemma.

Il requisito di partenza, quindi, è che tutti i documenti da organizzare in uno stemma siano codificati in XML. In particolare, ci si aspetta che ogni documento sia la trascrizione di un testimone di un'opera. Ci si aspetta, inoltre, che tutti i testimoni siano edizioni dello stesso tipo, quindi tutte edizioni diplomatiche o tutte edizioni interpretative elaborate secondo lo stesso insieme di regole.

Un punto da sottolineare è che non ci sono particolari vincoli nella scelta del vocabolario XML usato per la marcatura, l'unico obbligo è che in tutti i documenti venga usato lo stesso vocabolario. Vale la pena rimarcare però che la scelta del vocabolario può influenzare la qualità dello stemma generato: l'estrazione di differenze tra documenti basati su vocabolari più ricchi (come TEI) infatti tende ad essere più precisa e più utile durante la fase di ricerca degli errori significativi rispetto a vocabolari più poveri (come HTML).

Il nostro metodo, infatti, vuole sfruttare le informazioni strutturali e la classificazione degli errori in combinazione con gli strumenti di generazione automatica dello stemma, progettati invece per lavorare su contenuti puramente testuali.

8.3. Generazione automatica dello stemma

Nel corso degli anni sono stati sviluppati diversi strumenti, disponibili gratuitamente o a pagamento, per la generazione automatica dello stemma codicum. I più diffusi e sofisticati sono PAUP (Swofford 1993),

SplitsTree4 (Huson and Bryant 2006) e CompLearn (Rudi et al. 2008). Questi strumenti sono stati testati sia su documenti reali che su dati sintetici, anche attraverso l'organizzazione di competizioni pubbliche e workshop dedicati. Particolarmente rilevante è l'analisi in Roos and Heikkila 2009, che ha dimostrato l'applicabilità di questi metodi e la correttezza (almeno parziale) del loro output e ha evidenziato come il sistema PAUP produca i risultati migliori.

PAUP è un algoritmo per la ricostruzione filogenetica di alberi. Inizialmente sviluppato per la biologia evolutiva ha avuto notevoli applicazioni anche nel campo dell'ecologia, degli studi forensi e della filologia. A grandi linee, l'algoritmo prende in input un insieme di elementi, calcola le distanze tra le possibili coppie (usando diversi criteri di misurazione) e applica modelli matematici e calcoli probabilistici per derivare l'albero con alcune caratteristiche ottimizzate. La versione originale di PAUP, ad esempio, usa il metodo *parsimony*, cioè ricostruisce l'albero che minimizza il numero di cambiamenti tra coppie di elementi. Altre soluzioni integrano approcci diversi come *distance methods*, *likelihood method*, *recombination networks*, ecc. I dettagli matematici di queste soluzioni non sono rilevanti per la discussione ma è importante sottolineare l'approccio generale condiviso da questi algoritmi: sfruttare modelli matematici sofisticati e funzioni di ottimizzazione per derivare una possibile configurazione dell'albero evolutivo.

L'uso di questi algoritmi per la ricostruzione dello stemma codicum ne evidenzia alcune limitazioni, su cui è possibile lavorare per migliorare i risultati. Prima di tutto, il formato di input. Questi algoritmi infatti analizzano il puro contenuto testuale dei testimoni. Le informazioni strutturali non sono invece utilizzate affatto. La caratterizzazione delle varie parti di un testo potrebbe essere sfruttata per individuare gli errori (e soprattutto le motivazioni che li hanno generati) e fornire così informazioni utili per la costruzione dello stemma.

Tali algoritmi, inoltre, non sfruttano la classificazione degli errori. Essi infatti gestiscono tutti gli errori in modo molto simile: "appiattiscono" le differenze strutturali e semantiche a differenze tra sequenze di caratteri, su cui eseguono sofisticate analisi quantitative.

Un ulteriore problema deriva dal fatto che gli strumenti esistenti sono completamente automatizzati. Essi infatti prendono in input un

set di documenti collegati e producono un possibile output, spostando così l'intera logica di analisi all'interno dell'algoritmo e lasciando al filologo poco spazio di manovra nella personalizzazione del processo. Anche se supportano molti parametri di configurazione e permettono di raffinare i risultati, la complessità d'uso di questi strumenti è ancora molto elevata.

Con questi approcci c'è quindi un'implicita sovrapposizione tra due ruoli e due aree di competenza: il filologo, esperto di analisi dei testi e classificazione degli errori, e l'informatico, esperto di uso e configurazione degli strumenti di analisi.

8.4. Un metodo semi-automatico e dichiarativo

L'idea di base di questo articolo consiste nel sostituire gli approcci completamente automatici con un approccio semi-automatico, in cui il filologo può intervenire sui risultati continuando ad usufruire degli strumenti descritti nella sezione precedente, anche se in maniera indiretta. Il nostro obiettivo, infatti, non è presentare un nuovo algoritmo di generazione dello stemma ma un modello per nascondere la complessità degli strumenti esistenti e permettere al filologo di controllarne l'input e raffinarne l'output attraverso *un'interpretazione qualitativa* delle differenze.

Per esempio, in assenza di una configurazione specifica, tutti gli errori di uno stesso tipo (ad esempio, tutti gli omoteleuti) sono valutati come parimenti importanti; lo studioso può aumentare invece il *peso* di un certo errore (ad esempio, l'omoteleuto rispetto all'aplografia, o in modo ancora più fine l'omoteleuto che appare nella quarta pagina nel confronto tra il testimone β e il testimone γ) per rendere la presenza di quel certo fenomeno più rilevante ai fini dei calcoli che portano alla generazione dello stemma.

Un altro punto su cui bisogna porre l'accento è infatti l'impostazione "dichiarativa" dell'approccio proposto. Gli algoritmi esistenti usano un catalogo di errori sul quale basano i loro calcoli; questo catalogo è definito nel codice e non è modificabile, salvo la modifica di alcuni parametri di input.

Il metodo descritto in seguito, invece, prevede che il filologo classifichi gli errori e definisca regole per identificare questi errori a partire dalle differenze tra i testimoni. Sarà poi un'applicazione a

tradurre questa classificazione in parametri e pesi processati dagli strumenti di generazione dello stemma. Raffinare non vuol dire tuttavia cambiare manualmente i parametri degli algoritmi, ma usare strumenti che nascondono la loro complessità e si avvicinano al modo di ragionare e classificare le differenze proprio dei filologi.

Questo approccio iterativo permette al filologo di sperimentare: mentre l'uso degli algoritmi visti in precedenza prevede una singola esecuzione dell'algoritmo, in questo caso si può procedere per tentativi abbozzati e successive raffinement, che supportano in maniera più naturale il lavoro di studio e scoperta del filologo.

Per descrivere più in dettaglio il nostro metodo useremo un frammento dello Zibaldone di Giacomo Leopardi, codificato in XML e mostrato nella figura seguente.

```
<l>si spiega perché piaccia la luce del sole o della
luna, veduta in luogo dov'essi non</l>
<l>si vedano e non si scopra la sorgente della luce; un
luogo solamente in parte</l>
<l>illuminato da essa luce; il riflesso di detta
luce,</l>
<l>e i vari effetti materiali che ne derivano; il
penetrare di detta luce</l>
<l>in luoghi dov'ella divenga incerta e impedita, e non
bene si distingua, come</l>
<l>attraverso un canneto, in una selva, per li balconi
socchiusi ec. ec.;</l>
```

Fig. 8.1. Codifica in XML di un frammento dello Zibaldone di G. Leopardi.

Il nostro metodo consta di quattro fasi, spiegate in dettaglio nelle seguenti sotto-sezioni:

1. Ricerca delle differenze tra testimoni in formato XML
2. Catalogazione degli errori tramite pattern
3. Generazione dello stemma
4. Raffinement incrementali

8.4.1. Ricerca delle differenze tra testimoni in formato XML

Il primo passo è la generazione dei *delta* tra tutte le possibili coppie di documenti. I delta sono descrizioni delle differenze che esistono tra due documenti. Ogni delta $\Delta_{(S,T)}$ contiene una lista di cambiamenti che sono stati rintracciati tra il documento d'origine (identificato da S , *source*) e quello di destinazione (identificato da T , *target*).

Si consideri ad esempio il seguente testimone del frammento di esempio, in cui sono state evidenziate per chiarezza le differenze con il documento originale:

```
<l>si spiega perché piaccia la luce del sole o della
luna, veduta in luogo dov'essi non</l>
<l>si si vedano e non si scopra la sorgente della luce;
un luogo solamente in parte</l>
<l>illuminato da essa luce; il riflesso di detta
luce,</l>
<l>e i vari effetti materiali che ne derivano; il
penetrare di detta luce</l>
<l>in luoghi dov'ella divenga incerta e impedita, e non
bene si distingua, come</l>
<l>attraverso un canneto, in una selva, per li balconi
socchiusi ec. ec.;</l>
```

Fig. 8.2. Testimone B.

Un esempio di delta è illustrato in figura 8.3, in cui si ritrovano due cambiamenti: l'aggiunta del testo "si" (cambiamento c_1) e la cancellazione della quarta linea (c_2 , cancellazione di un nodo XML con tutto il suo contenuto). I dettagli sintattici non sono rilevanti in questo contesto ma l'insieme di cambiamenti riconosciuti pone le basi per le successive analisi.

```
c1: ADD-TEXT("si ", /l[2])
c2: DEL-XML-SUBTREE(/l[4])
```

Fig. 8.3. Delta tra A e B.

Esistono diversi strumenti automatici per calcolare il delta a partire da documenti XML, tra cui ricordiamo XyDiff (Cóbena et al. 2002), Faxma (Lindholm et al. 2006) e DocTreeDiff (Rönnau et al. 2009). Alcuni tra questi sono progettati cercando di ottimizzare i costi computazionali e produrre il delta velocemente, altri per minimizzare lo spazio occupato dal delta, altri per ridurre al minimo l'elenco delle operazioni inserite nel delta. Essi differiscono inoltre per l'insieme di operazioni che sono in grado di riconoscere (ad esempio, solo alcuni riconoscono lo spostamento dei contenuti) e per la granularità e precisione delle differenze trovate. Quest'ultimo punto è fondamentale per la buona riuscita del nostro approccio: migliore è la qualità dell'output, infatti, maggiori saranno le informazioni disponibili nei passi successivi. In particolare, è fondamentale che l'algoritmo sappia esprimere fedelmente sia i cambiamenti al testo, sia quelli alla struttura e sia quelli combinati.

Stiamo lavorando, infatti, sia agli strumenti di verifica della qualità di questi algoritmi e sia ad una specifica implementazione particolarmente adatta a questo contesto. JNDiff (Di Iorio et al. 2009) è un'implementazione open-source, che riconosce un vasto numero di operazioni e produce ottimi risultati sui documenti letterari. Caratteristica principale dell'algoritmo è la possibilità di aggregare cambiamenti atomici (inserimenti, cancellazioni) in cambiamenti complessi e più significativi per l'utente (es. spostamento di paragrafo). Una delle direzioni di ricerca su JNDiff è lo studio di un modulo per riconoscere cambiamenti significativi sui testimoni e produrre così un output più utile in questa fase in quanto più vicino alla tipologia di errori rilevanti in campo filologico.

8.4.2. Catalogazione degli errori tramite pattern

Il passo successivo consiste nell'analisi dei delta e nell'interpretazione dei cambiamenti per ottenere un elenco di errori significativi. Il termine "significativo" in questo caso indica "errori espressi secondo la terminologia e lo schema mentale noto al filologo". L'obiettivo è individuare omeoarchie, dittografie, errori per omoteleuti, errori di ripetizione, errori di anticipo, e così via. Studiando questi errori e le loro possibili cause ed effetti, il filologo può ricostruire le relazioni tra documenti e lo stemma codicum. È qui che l'esperienza e l'abilità entrano in gioco, ed è quindi interessante permettergli di studiare e

classificare gli errori a partire dall'output prodotto dagli algoritmi visti nella precedente sezioni.

L'idea di base, infatti, è sintetizzare errori significativi a partire dai cambiamenti ritrovati all'interno dei delta, attraverso un meccanismo di *pattern matching*. I pattern descrivono quali sono le condizioni per riconoscere un certo tipo di errore. In particolare, descrivono come la presenza di alcuni cambiamenti atomici nel delta e di alcune caratteristiche del testo o della struttura modificata, possono essere tradotti in cambiamenti significativi.

Si consideri la struttura XML del frammento di esempio e il delta mostrato in figura 8.3. Il cambiamento c_2 è un errore per omoteleuto. Esso può essere individuato se:

- nel delta $\Delta_{(S,T)}$ è presente un cambiamento con operazione DEL-XML-SUBTREE che incide sul sotto-albero XML che ha come radice l'elemento XML e ;
- l'elemento XML e è un elemento di tipo TEI <I>.
- il contenuto testuale rimosso in c si conclude con una parte di testo identica alla parte finale del testo contenuta nel sottoalbero con radice e_{prev} , ovvero con la parte finale dell'elemento <I> precedente.

Questa definizione può essere descritta in un pattern formale come quello mostrato in figura 8.4.

OMOTELEUTO:

```
exists DEL-XML-SUBTREE(c)

elem_pos = c.deleted_element.pos
equals c.deleted_element.type, "TEI:1"
equals Source.element_at(elem_pos-1).type, "TEI:1"

end1 = last-characters(c.text)
end2 = last-characters(Source.text_at(elem_pos-1))
equals end1, end2
```

Fig. 8.4. Pattern per il riconoscimento di omoteleuti.

Indipendentemente dagli aspetti sintattici, usati solo per la spiegazione e non legati ad una particolare implementazione, è importante sottolineare il ruolo dei pattern come meccanismo di astrazione dei cambiamenti. Si noti inoltre che, una volta definito il pattern, la fase di riconoscimento può essere completamente automatizzata. Il compito del filologo consiste, quindi, nella creazione di un catalogo di pattern che sarà poi processato da uno strumento automatico in fase di matching. Questo schema permette anche di raffinare la definizione di un pattern e iterare l'analisi con pochi sforzi.

Da notare che non ci aspettiamo che il filologo scriva direttamente la definizione formale del pattern seguendo la sintassi vista nell'esempio (o simili) ma prevediamo l'esistenza di strumenti semplificati di creazione e affinamento delle regole. Un punto collegato è che il catalogo può essere personale ma anche condiviso da un gruppo o da una comunità più ampia. In questo modo si evita di dover definire più volte gli stessi pattern e si rendono possibili confronti incrociati.

Il catalogo dei pattern è quindi aperto e l'approccio è indipendente da una specifica classificazione degli errori, come ad esempio quella proposta da Dain 1964 o più recentemente da Conti 2011.

Si consideri ora un secondo testimone mostrato in figura 8.5.

```

<l>si spiega perché piaccia la luce del sole o della luna,
veduta in luogo dov'essi non</l>
<l>si vedano e non si scopra la sorgente della luce; un
luogo solamente in parte</l>
<l>illuminato da essa luce; il riflesso di detta luce,</l>
<l>e i vari effetti materiali che ne derivano; il penetrare
di detta luce</l>
<l>in in luoghi dov'ella divenga incertata e impedita, e
non benemale si distingua, come</l>
<l>attraverso un canneto, in una selva, per li balconi
socchiusi ec. ec.;</l>

```

Fig. 8.5. Testimone C. Per chiarezza sono evidenziate le differenze con il documento originale A.

Confrontando il testimone C con l'archetipo A (figura 8.1) un algoritmo per il calcolo delle differenze costruisce il seguente delta:

```

c1: ADD-TEXT("in ", /l[4]/text()[4])
c2: ADD-TEXT("ta ", l[4]/text()[30])
c3: DEL-TEXT("male", /l[4]/text()[51])
c4: ADD-TEXT("bene", /l[4]/text()[51])

```

Fig. 8.6. Delta tra A e C.

Un possibile pattern in grado di riconoscere una dittografia su questo delta verifica se, in presenza di un'aggiunta di caratteri, il frammento di testo immediatamente successivo o immediatamente precedente contiene gli stessi caratteri (o una parte di essi). Nel frammento di esempio vuol dire verificare se il nodo di testo individuato con `/l[4]/text()[4]` nel sorgente contiene la sequenza "in " oppure se il nodo di testo individuato con `/l[4]/text()[30]` nel sorgente contiene la sequenza "ta ". Il processo di pattern matching porterebbe quindi a raffinare il delta in:

```

c1: DITTOGRAFIA("in ", /l[4]/text()[4])
c2: DITTOGRAFIA("ta ", l[4]/text()[30])
c3: DEL-TEXT("male", /l[4]/text()[51])
c4: ADD-TEXT("bene", /l[4]/text()[51])

```

Fig. 8.7. Delta tra A e C, con l'individuazione di dittografie.

Si noti inoltre che, sfruttando un modulo in grado di riconoscere che i due termini "male" e "bene" rispettivamente cancellati ed inseriti nella stessa posizione hanno significato opposto, si potrebbe raffinare ulteriormente questo delta e individuare un errore polare, come mostrato in figura 8.8.

```

c1: DITTOGRAFIA("in ", /l[4]/text()[4])
c2: DITTOGRAFIA("ta ", l[4]/text()[30])
c3: ERRORE-POLARE("male", "bene", /l[4]/text()[51])

```

Fig. 8.8. Delta tra A e C, con l'individuazione di cambiamenti significativi.

Si consideri infine un quarto testimone mostrato in figura 8.9.

```

<l>si spiega perché piaccia la luce del sole o della
luna, veduta in luogo dov'essi non</l>
<l>si vedano e non si scopra la sorgente della luce; un
luogo solamente in parte</l>
<l>illuminato da essa luce; il riflesso di detta
luce,</l>
<l>e i vari effetti materiali che ne derivano; il
penetrare di detta luce</l>
<l>in in luoghi dov'ella divenga incerta e impedita, e
non si distingua, come</l>
<l>attraverso un canneto, in una selva, per li balconi
socchiusi ec. ec.;</l>

```

Fig. 8.9. Testimone D.

Ci sono due differenze tra questo testimone e l'archetipo in figura 8.1: ripetizione dei caratteri "in" e cancellazione del testo "bene", entrambi alle linea 4. Eseguendo un algoritmo per il calcolo delle differenze ed applicando il pattern per riconoscere le dittografie si ottiene il risultato mostrato in figura 8.10.

```

c1: DITTOGRAFIA("in ", /l[4]/text()[4])
c2: DEL-TEXT("bene", /l[4]/text()[51])

```

Fig. 8.10. Delta tra A e D.

Confrontando questo stesso testimone con il testimone C in figura 8.5, invece, l'algoritmo costruisce il delta mostrato in figura 8.11, su cui non è possibile alcuna interpretazione con i pattern visti finora.

```
c1: DEL-TEXT("ta", /l[4]/text()[30])
c2: DEL-TEXT("male", /l[5]/text()[4])
```

Fig. 8.11. Delta tra C e D.

L'ultimo significativo passo di questa fase è l'associazione di *pesi* ai vari tipi di errori o errori specifici. Questi pesi infatti permettono di tradurre l'analisi sugli errori in informazioni quantitative, a loro volta usate dagli algoritmi per la generazione dello stemma.

L'idea è usare valori che variano tra 0 e 1, dove 0 indica la minima importanza e 1 la massima. I pesi sono associabili agli errori a vari livelli. Innanzitutto si possono associare dei pesi generici ad ogni tipo di errore, ad esempio dicendo che tutte le dittografie hanno come peso 0,2, gli omoteleuti 0,5 e gli errori polari 0,9. È poi possibile indicare che nel confronto di una certa coppia di testimoni le dittografie hanno valore 0,6, quindi più importanti degli omoteleuti ma solo per i confronti fatti su quella specifica coppia. Infine è possibile indicare il peso di un preciso errore, ad esempio impostare a 0,01 il peso di un certo errore polare riscontrato in un certo confronto perché si sa da fonti esterne che tale errore è di importanza minima.

Il calcolo dei pesi permette di associare valori numerici alle coppie di testimoni usati in seguito per generare lo stemma. Si noti che nell'esempio tratto dallo Zibaldone non sono stati analizzati tutti i delta per questioni di sinteticità e chiarezza.

8.4.3. Generazione dello stemma

La generazione vera e propria della stemma è realizzata usando strumenti già noti in letteratura, come quelli descritti nella sezione 3. Come osservato in precedenza, gli algoritmi non vengono utilizzati direttamente sulla collezione di documenti ma su dati che sono sintetizzati sulla base degli errori trovati e dei pesi a loro associati. È compito dell'esperto di informatica implementare il modulo di traduzione di un delta con errori significativi in una serie di input numerici per gli strumenti automatici, con l'obiettivo di nascondere tale complessità al filologo. Si noti anche che questo modello è indipendente da uno specifico algoritmo di generazione dello stemma.

Laddove è possibile implementare i meccanismi di traduzione, sarà

inoltre possibile sperimentare diversi strumenti sugli stessi dati di input.

In questa fase non abbiamo ancora implementato meccanismi di traduzione e non ancora testato tali strumenti su documenti reali. È possibile, tuttavia, simulare lo stemma generato per gli esempi visti finora, come mostrato in figura 8.12.

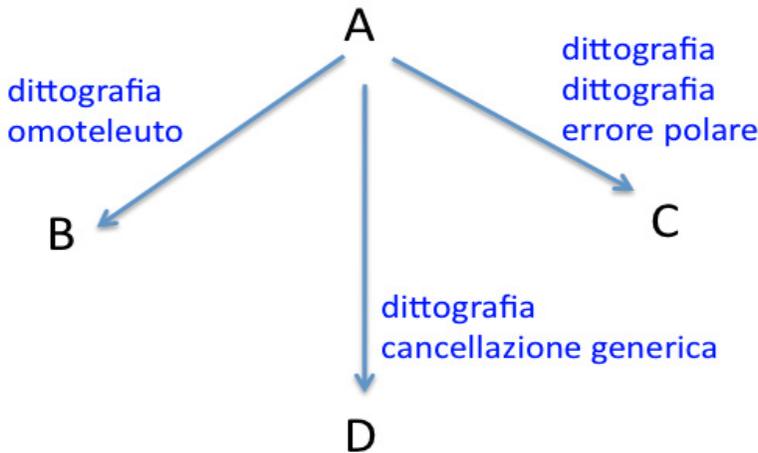


Fig. 8.12. Possibile stemma dei testimoni in esempio.

8.4.4. Raffinamenti incrementali

La generazione dello stemma codicum non si esaurisce però con la creazione della prima soluzione. È lecito aspettarsi che lo studioso voglia apportare delle modifiche prima di essere soddisfatto del risultato ottenuto. Ci sono diversi aspetti attraverso i quali è possibile modificare il risultato, sia per integrare nuove conoscenze sopraggiunte che per sperimentare varie ipotesi di lavoro.

Il tipo di raffinamento più comune è il cambio dei pesi associati ad alcuni tipi di errori oppure a singoli cambiamenti, descritto in precedenza. Poiché la generazione meccanica dello stemma è completamente automatica, cambiare i pesi e reiterare il processo non è dispendioso per il filologo.

Un altro tipo di raffinamento possibile è l'aggiunta al catalogo di nuovi errori o la modifica dei pattern esistenti. Ad una prima iterazione, infatti, il confronto tra due testimoni potrebbe individuare cambiamenti non associabili a nessun pattern. Ad esempio, nel delta ricostruito tra i testimoni C e D e mostrato in figura 8.11 è individuata una cancellazione generica della sillaba "ta". Tale cancellazione può essere invece classificata come aplografia e identificata da un pattern. Aggiungendo il pattern al catalogo e eseguendo di nuovo l'analisi il delta può essere raffinato in:

```
c1: APLOGRAFIA("incertata", "incerta", /l[4]/text()[26])
c2: DEL-TEXT("male", /l[5]/text()[4])
```

Fig. 8.13. Delta tra i testimoni C e D, con riconoscimento esplicito di un'aplografia.

Questo delta sarà tradotto in una nuova configurazione in input per l'algoritmo di generazione dello stemma che, nel caso in cui l'aplografia abbia un'alta rilevanza (peso maggiore), potrebbe cambiare come mostrato in figura 8.14.

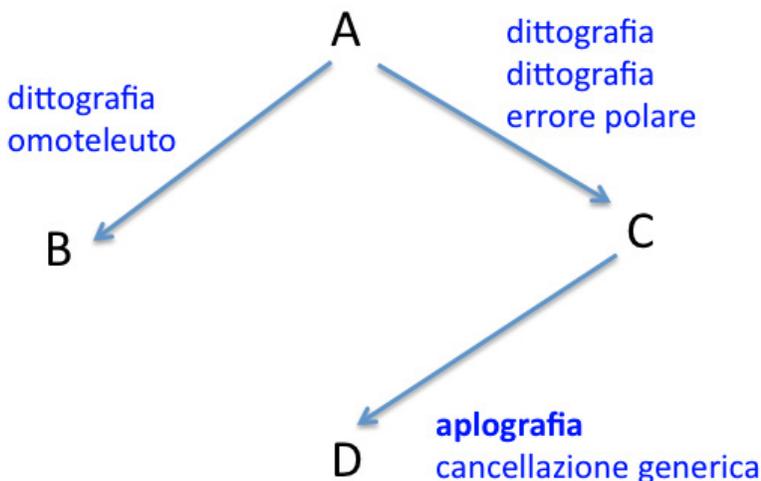


Fig. 8.14. Stemma alternativo, in presenza di un'aplografia con peso maggiore.

Si noti infine che il nostro metodo prevede anche la possibilità di forzare una certa relazione tra due testimoni (specificando pesi globali sulla coppia) in modo da coprire anche i casi in cui, ad esempio per conoscenza da fonti esterne, alcuni legami nello stemma non sono possibili o sono accertati.

8.5. Conclusioni

L'approccio presentato in questo articolo si basa su una netta separazione tra l'analisi filologica di un testo e le operazioni meccaniche a supporto di questa analisi, con l'idea di demandare queste operazioni a sistemi software e permettere al filologo di concentrarsi sulla classificazione degli errori, riconoscimento dei pattern e mappatura con effetti sullo stemma.

L'obiettivo della ricerca, infatti, è la realizzazione di un ambiente di lavoro che lo studioso di filologia possa usare per sperimentare in fase di analisi e generazione dello stemma.

L'uso della codifica XML e degli algoritmi per il calcolo delle differenze in questo contesto può essere infine interessante sia per gli studiosi di filologia (che hanno la possibilità di approfondire la classificazione degli errori, l'analisi dei pattern di editing e delle relazioni tra testimoni) e sia per gli esperti di informatica (che possono sperimentare un approccio domain-oriented agli algoritmi di diff, meccanismi di verifica della qualità dei delta e nuove tecniche di pattern matching).

8.6. Bibliografia

- AIDAN CONTI. Wherefore a typology of variants. In *6th International Workshops on Stemmatology*, Helsinki, Finland, 2011.
- GRÉGORY CÔBENA, SERGE ABITEBOUL, AND AMÉLIE MARIAN. Detecting changes in XML documents. In *Proceedings of the 18th International Conference on Data Engineering*, pages 41–52, San Jose, CA, USA, 2002.
- A. DAIN. *Les manuscrits*. Collection d'études anciennes. Les Belles lettres, 1964.

- STEVEN J. DEROSE. Xml and the tei. *Computers and the Humanities*, 33 (1-2):11–30, 1999.
- ANGELO DI IORIO, MICHELE SCHIRINZI, FABIO VITALI, AND CARLO MARCHETTI. A natural and multi-layered approach to detect changes in tree-based textual documents. In *Proceedings of 11th International Conference Enterprise Information Systems, ICEIS 2009*, pages 90–101, Milan, Italy, 2009.
- DANIEL H HUSON AND DAVID BRYANT. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, 23(2):254–267, February 2006. ISSN 0737-4038 (Print). doi: 10.1093/molbev/msj030.
- TANCRED LINDHOLM, JAAKKO KANGASHARJU, AND SASU TARKOMA. Fast and simple XML tree differencing by sequence alignment. In *ACM Symposium on Document Engineering*, pages 75–84, 2006.
- TEEMU ROOS AND TUOMAS HEIKKILA. Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets. *Literary and Linguistic Computing*, 24 (4):417–433, 2009. doi: 10.1093/lc/fqp002.
- STEVEN DE ROOIJ RUDI CILIBRASI, ANNA LISSA CRUZ AND MAARTEN KEIJZER. Complearn, 2008. URL <http://www.complearn.org/>.
- SEBASTIAN RÖNNAU, GERAINT PHILIPP, AND UWE M. BORGHOFF. Efficient change control of XML documents. In *ACM Symposium on Document Engineering*, pages 3–12, 2009.
- DAVID SWOFFORD. Paup: Phylogenetic analysis using parsimony, 1993. URL <http://paup.csit.fsu.edu/>.

9. Il progetto Visionary Cross: verso un'edizione digitale multimediale e distribuita

Roberto Rosselli Del Turco¹

Abstract. Il progetto Visionary Cross ha come obiettivo un'edizione multimediale avanzata di testi e monumenti dell'era anglosassone: *Elena* e *Il Sogno della Croce*, due poemetti in inglese antico (*Codex Vercellensis*, segnatura MS CXVII, Archivio e Biblioteca Capitolare di Vercelli); le croci monumentali di Ruthwell (Dumfriesshire, Scotland) e Bewcastle (Cumbria, England); e la croce di Bruxelles (Cattedrale dei SS. Michele e Gudula, Bruxelles, Belgio). Il tema che unisce tutti questi manufatti è quello del culto della Croce e, in particolare, la Leggenda della Vera Croce (*inventio crucis*), assai significativo in epoca medievale e comune a molte nazioni europee. La definizione dell'edizione richiede una riflessione teorica su metodi e tecniche della filologia digitale, come pure la progettazione di una interfaccia utente in grado di visualizzare e navigare le complesse interrelazioni fra gli oggetti che ne fanno parte. Il risultato finale sarà basato su un *framework* aperto e collaborativo.

Parole chiave: inglese antico, letteratura anglosassone, storia dell'arte anglosassone, filologia digitale, modellazione 3D.

¹ Dipartimento di Studi Umanistici, Università di Torino.
e-mail: roberto.rosselidelturco@unito.it.

9.1. Il progetto *Visionary Cross*

Il progetto *Visionary Cross*² nasce nella primavera del 2005, quando in occasione del convegno annuale di studi medievali di Kalamazoo³, e in maniera un po' fortuita, si trovarono insieme a discutere di cose anglosassoni alcuni studiosi, in particolare Catherine Karkov, Daniel Paul O'Donnell, Martin Foys e il sottoscritto. Da una presentazione di M. Foys relativa alla croce di Ruthwell, nella quale grazie a un'applicazione Flash un modello 3D veniva 'disassemblato' in maniera da mostrare le parti non originali della croce e la posizione corretta della parte apicale, era nata l'idea di creare un'edizione integrata sia della *standing cross* anglosassone, sia del *Sogno della Croce*: il poemetto in inglese antico conservato nel Vercelli Book è infatti strettamente connesso alla croce di Ruthwell in quanto quest'ultima riporta alcuni versi del primo incisi nella pietra con caratteri runici. Da allora sono successe molte cose, il progetto è rimasto in uno stato 'dormiente' mentre cercavamo di definirne meglio gli aspetti fondamentali (e di reperire i fondi necessari!) e ha perso uno dei fondatori (M. Foys), mentre per motivi che saranno chiariti nel paragrafo successivo è stato ampliato il focus ad altre due croci (croce di Bewcastle e croce di Bruxelles) e al poema *Elena*, anche quest'ultimo tradito dal Vercelli Book; resta per il momento fuori dalla lista il testo omiletico in inglese antico *The Finding of the True Cross*, ma sarà sicuramente preso in considerazione per una futura espansione dell'edizione.

Nel corso di questa ridefinizione e ampliamento del numero di oggetti che entreranno a far parte dell'edizione, tuttavia, l'impostazione originale basata su un 'dialogo' a distanza fra testi presenti su supporti di natura radicalmente diversa si è ben presto rivelata insufficiente sia sul piano metodologico, sia anche per quanto riguarda gli obiettivi stessi del progetto. L'inclusione della croce di Bewcastle e di quella di Bruxelles è infatti dovuta all'individuazione di numerosi e importanti collegamenti, non solo di tipo testuale, fra questi manufatti e i diversi supporti del testo del *Sogno della Croce*: la funzione di *preaching cross*,

² Si veda il sito <http://www.visionarycross.org/> per ulteriori informazioni e per un aggiornamento sugli sviluppi del progetto.

³ International Congress on Medieval Studies: <http://www.wmich.edu/medieval/congress/>.

alcuni pannelli figurativi con lo stesso soggetto, i motivi decorativi di origine celtica e le iscrizioni runiche per Ruthwell e Bewcastle; il testo del *Sogno della Croce* inciso su Ruthwell e Bruxelles; l'iscrizione di dedica per Bewcastle e Bruxelles. *L'Elena* è stata inclusa non soltanto perché è nel Vercelli Book come il *Sogno*, ma anche perché condivide con quest'ultimo i temi del culto della Vera Croce e delle visioni ad Essa collegate. Il significato di tali manufatti dipende in larga parte dalla complessa interazione fra testo, immagini e contesto: siamo pertanto arrivati alla conclusione che limitare l'edizione a una presentazione di modelli tridimensionali con qualche collegamento al testo delle edizioni critiche avrebbe avuto poco senso e sarebbe stato uno spreco di risorse⁴.

Un'altra considerazione emersa nel corso delle discussioni fra i partecipanti al progetto riguarda i più recenti sviluppi nel campo delle edizioni online. Due nuove tendenze, in particolare, sono apparse promettenti: la possibilità di interconnessione fra edizioni ospitate da siti web indipendenti e il cosiddetto 'aspetto sociale' delle edizioni sul web. Nel primo caso, il punto di riferimento è l'iniziativa NINES (*Networked Infrastructure for Nineteenth-Century Electronic Scholarship*), una organizzazione accademica "devoted to forging links between the material archive of the nineteenth century and the digital research environment of the twenty-first. [...] The NINES Collex interface⁵ is at the center of these efforts. It aims to gather the best scholarly resources in the field and make them fully searchable and interoperable; and to provide an online collecting and authoring space in which researchers can create and publish their own work."⁶ Questo principio di interoperabilità e la possibilità di effettuare ricerche contemporaneamente su tutti i siti affiliati sono stati raccolti da MESA (*Medieval Electronic Scholarly Alliance*)⁷, progetto di recentissima costituzione che

⁴ Dopo una fase di entusiasmo iniziale, dovuto alla novità dei modelli 3D di manufatti di vario tipo e alle nuove possibilità di ricerca che questi rendono possibili, la semplice presenza di uno o più modelli di questo tipo non è considerata, di per sé, elemento di innovazione: è essenziale inserire il modello in un contesto informativo ed editoriale completo, e fornire strumenti adeguati, testuali e non testuali, per la ricerca.

⁵ <http://www.nines.org/about/software/collex/>.

⁶ Si veda il sito <http://www.nines.org/>.

⁷ Ulteriori informazioni disponibili all'URL <http://www.dlib.indiana.edu/projects/mesa/>.

ambisce a svolgere la medesima funzione di NINES nell'ambito degli studi di medievistica. La partecipazione di Visionary Cross al network MESA è stata quindi considerata prioritaria, e per tale ragione particolare attenzione verrà data alla produzione di metadati relativi ai differenti oggetti inseriti nel *framework* dell'edizione finale; questa strategia garantisce anche la possibilità di partecipare ad altre reti di biblioteche digitali, come ad esempio Europeaana.

L'altro fenomeno ritenuto pertinente ai fini del progetto, strettamente correlato alla nascita e diffusione del cosiddetto "web 2.0", è l'introduzione dell'aspetto collaborativo e distribuito nelle edizioni digitali (Siemens 2012). Si tratta di una caratteristica inedita, importante perché aggiunge un livello di contatto, interazione e collaborazione con i fruitori delle edizioni digitali del tutto impossibile con mezzi tradizionali, ma anche potenzialmente pericolosa, per lo meno in base al livello di interazione prescelto: una eventuale compartecipazione degli utenti al livello di responsabilità più alto, quello editoriale, richiederebbe massima cautela e una buona quantità di filtri.

Si può pertanto affermare che lo scopo del progetto non è soltanto quello di proporre nuove edizioni di alcuni testi e manufatti di fondamentale importanza per gli studi anglosassoni, ma soprattutto di utilizzare un certo numero di nuove tecnologie e metodologie, da poco emerse come particolarmente significative nel campo dell'Informatica Umanistica, per produrre un'edizione innovativa, con il duplice obiettivo di contribuire in maniera significativa al progresso sia degli studi anglosassoni, sia delle stesse metodologie impiegate, e di creare una infrastruttura aperta, espandibile nell'ambito del progetto e riutilizzabile per altri progetti, in cui la combinazione di robustezza e validità del contenuto editoriale, dinamicità nella gestione dei singoli oggetti, aspetto collaborativo dell'edizione permetteranno di raggiungere risultati impossibili da conseguire con una tradizionale edizione critica a stampa.

9.2. Monumenti e componenti⁸

9.2.1. Il Vercelli Book

Il Codex Vercellensis, meglio noto come Vercelli Book, conservato a Vercelli presso l'Archivio e Biblioteca Capitolare di Vercelli con segnatura cod. CXVII, è un manoscritto databile alla seconda metà del X secolo il cui luogo di composizione è tuttora ignoto, ma sicuramente riconducibile a uno *scriptorium* dell'area sud-orientale dell'Inghilterra. Redatto da un'unica mano in grafia minuscola quadrata anglosassone nel dialetto tardo sassone occidentale, questo codice è particolarmente rilevante per gli studi di lingua e letteratura in inglese antico perché appartiene a un esiguo drappello di manoscritti (solo quattro!) che conservano la maggior parte dei testi poetici in inglese antico⁹. Malgrado si tratti, quindi, di un documento fondamentale per tutti gli studiosi del settore, è di aspetto sorprendentemente semplice, sostanzialmente privo di iniziali miniate¹⁰ o di altre forme di decorazione¹¹, forse perché si trattava di un *ascetic florilegium* per uso privato, come suggerito da Éamonn Ó Carrágain (Ó Carrágain 1975): l'ipotesi di un manoscritto 'da viaggio', per stimolare la riflessione e la preghiera di

⁸ La bibliografia relativa ai componenti poetici e ai manufatti artistici oggetto della nostra ricerca è molto ampia, pertanto in questa nota mi limiterò a dare qualche indicazione bibliografica fondamentale in aggiunta ai testi citati nel testo dell'articolo; altri riferimenti utili sono presenti nella bibliografia finale. Per quanto riguarda il Vercelli Book si veda Förster 1913 e Sisam 1976 per una minuziosa descrizione paleografica; in italiano, Luiselli Fadda 1994 per un inquadramento generale del manoscritto nella tradizione anglosassone, Scragg 1973 e Ó Carrágain 1975 in merito alla sua compilazione; ottimo per quanto riguarda le tendenze più recenti della critica Zacher 2009; Swanton 1970 e Gradon 1977 per le edizioni critiche del Sogno della Croce e di Elena. Sul tema della inventio crucis Drijvers 1997, per la tradizione poetica anglosassone Luiselli Fadda 2000 e Ó Carrágain 2005; per quanto riguarda le croci di Ruthwell e Bewcastle dal punto di vista della storia dell'arte anglosassone si vedano, tra gli altri, Karvov 2003b e Brown 2008; per la croce di Buxelles D'Ardenne 1939 e il già citato Ó Carrágain 2005.

⁹ Gli altri tre codici sono l'Exeter Book, il Cotton Vitellius A. XV (che conserva il poema *Beowulf*) e lo Junius XI.

¹⁰ Con l'eccezione di tre iniziali maiuscole decorate con motivi zoomorfici ai ff. 49r (H), 106r e 112r (M in entrambi i casi).

¹¹ Sono presenti pochissime decorazioni, come una figura zoomorfica nel margine inferiore del f. 49v e un'altra, limitata a una testa di animale e di fattura assai più rozza, al f. 112r. Lo scriba, inoltre, ha usato inchiostri colorati con estrema parsimonia, ad esempio al f. 90v (giallo chiaro) e ai ff. 71v e 73v (rosso).

un pio pellegrino, ci permette forse di capire perché il Vercelli Book, a differenza degli altri manoscritti contenenti testi poetici in inglese antico, sia conservato in Italia e non in Inghilterra: molte delle ipotesi riguardo il suo arrivo a Vercelli, una delle tappe fondamentali della via Francigena, partono infatti dal presupposto che il manoscritto appartenesse a un religioso (forse addirittura a un vescovo) anglosassone che aveva intrapreso un viaggio, forse un pellegrinaggio, nell'Europa continentale.

Il codice consta di 135 fogli pergamenei, molto sottili, suddivisi in 19 fascicoli, e contiene ventitré omelie intervallate, apparentemente senza uno schema preciso, da sei componimenti poetici: l'*Andreas* (ff. 29v-52v), i *Fati degli Apostoli* (ff. 52v-54r), il *Discorso dell'anima al corpo* (ff. 101v-103v), il *Frammento Omiletico I* (ff. 104r-104v), il *sogno della Croce* (ff. 104v-106r) e l'*Elena* (ff. 121r-133v). La disposizione dei fascicoli e dei testi ivi contenuti, come pure la diversa spaziatura delle righe di testo a partire dal XV fascicolo e altri fenomeni secondari, ad esempio la variazione delle dimensioni della grafia, ridotte in alcuni punti, lasciano supporre che sia stato composto in un arco di tempo relativamente lungo, unendo fascicoli redatti in momenti diversi.

9.2.2. *Elena*

Questo componimento di 1321 versi è uno dei pochi di epoca anglosassone che può essere attribuito a un autore con sicurezza grazie alla sua 'firma' runica: un acrostico del nome Cynewulf composto usando caratteri runici (f. 133r). L'*Elena*, composta nella prima metà del IX secolo, è la prima versione in inglese della leggenda relativa al ritrovamento della Croce (*inventio crucis*) da parte di S. Elena, madre dell'imperatore Costantino, basata probabilmente su una versione in latino del testo dell'*Acta Syriaci*, una delle versioni della leggenda più diffuse (Nestle 1889). La visione della Croce da parte di Costantino riportata nella parte iniziale collega questo componimento al *Sogno della Croce*, e quindi anche alla croce di Ruthwell: che questi due temi leggendari (ritrovamento della Vera Croce e visione di Costantino) siano spesso intrecciati e connessi nella letteratura religiosa del periodo lo dimostra anche il fatto che, in alcune versioni della prima, il luogo in cui è celata la Croce viene rivelato a Elena per mezzo di una visione. Se la narrativa segue la leggenda dell'*inventio crucis* in maniera abbastanza fedele, l'autore mostra spesso una notevole originalità nel

trattare l'argomento: ad esempio Elena è descritta come una donna decisa e risoluta, una vera eroina cristiana determinata a concludere con successo la sua sacra missione.

9.2.3. Il sogno della Croce

Il *Sogno della Croce*, uno dei componimenti più affascinanti di tutto il corpus della letteratura in inglese antico, il *Sogno della Croce* inizia con la descrizione di una visione portentosa comparsa in sogno al poeta, un'immagine sovranaturale che assume via via forme diverse: un albero, un simbolo luminoso e, infine, una croce; anche questo simbolo si dimostra assai mutevole, aparendo a volte coperta dal sangue e dai segni delle torture subite dal Figlio di Dio, altre (come la croce di Bruxelles) da oro e gioielli. La Croce narra quindi in prima persona la propria storia, e il poemetto termina con le istruzioni impartite dalla Croce al poeta, incoraggiato a raccontare la propria visione in modo che tutta l'umanità possa ascoltare il messaggio di Cristo e salvarsi. Se l'iscrizione runica di Ruthwell è coeva con il resto del monumento, come tutto lascia supporre, allora questo poemetto ha una storia testuale che è più lunga e più diversificata, sul piano linguistico e geografico, di quasi ogni altro componimento in volgare del periodo anglosassone: non solo, infatti, si tratta di uno dei pochissimi testi letterari in inglese antico a non essere stato trasmesso come *textus unicus*, ma il fatto che l'unico altro frammento di una certa lunghezza (la croce di Bruxelles riporta solo due versi) sia stato trasmesso come iscrizione epigrafica in caratteri runici su di una *preaching cross* più antica di almeno due secoli rispetto al manoscritto che ci tramanda il poemetto nella sua interezza (cfr. il paragrafo successivo) rende il suo caso del tutto eccezionale.

9.2.4. La croce di Ruthwell

La croce di Ruthwell è una croce monumentale in pietra, alta circa 5,28 metri, su pianta rettangolare con tendenza a rastremarsi verso l'alto (circa 72x53,8 cm alla base, 33x24 cm alla sommità) che si trova nel Dumfriesshire (Scozia). Eretta nel corso dell'VIII secolo in quello che era un antico sito fortificato romano, in seguito trasformato in un monastero, la croce restò indisturbata per secoli finché nel 1642, come conseguenza della Riforma, venne promulgato l'editto 'Idoltrous monuments in Ruthwell' che ne decretò l'abbattimento: la furia iconoclasta fu tale che la croce venne spezzata in più frammenti, alcuni

dei quali addirittura danneggiati nel tentativo di obliterare per sempre quanto raffigurato (è il caso della scena della crocifissione, alla base del lato sud, che mostra ancora visibili i segni lasciati dagli scalpelli); alcuni dei frammenti furono impiegati come pavimentazione della vicina chiesa, altri vennero sepolti nel terreno del cimitero, e sono stati recuperati solo in parte. Nel 1802 il rev. Henry Duncan fece ricostruire la croce, inizialmente in una posizione all'esterno della chiesa di Ruthwell, e nel 1823 il braccio orizzontale, del quale era stato ritrovato solo un frammento, fu sostituito con una replica con decorazioni moderne. In un momento successivo, nel 1887, la croce fu spostata in un'abside creata appositamente per questo scopo. Questo monumento è molto conosciuto fra gli studiosi di area anglosassone anche in virtù dell'iscrizione in caratteri runici che conserva un frammento del poemetto *Il sogno della Croce* contenuto nel Vercelli Book, e che potrebbe essere la più antica conosciuta di un componimento poetico in inglese antico. Oltre a questi versi composti secondo il metro tradizionale anglosassone, la croce di Ruthwell presenta decorazioni di ispirazione celtica sui lati più stretti e scene ispirate alle Scritture, accompagnate da iscrizioni in latino, su quelli più larghi.

9.2.5. La croce di Bewcastle

Anche la croce di Bewcastle, alta circa 4,5 m. e con base quadrata più che rettangolare (56x54 cm), sorge sul sito di un insediamento fortificato di origine romana, successivamente convertito in comunità religiosa, e risale approssimativamente allo stesso periodo: "Although there is lively discussion about the dates of these monuments, there is a growing consensus that both are to be dated to the first half of the eighth century: as it were, to the "Age of Bede" (who died in 735) or to the generation after his death." (Ó Carragáin 2007). Essa presenta molti elementi in comune con la croce di Ruthwell: una serie di pannelli figurativi, alcuni dei quali raffigurano la medesima scena di alcuni tra quelli presenti su quest'ultima: è il caso di Cristo trionfante sulle bestie (o Cristo riconosciuto come Giudice dalle bestie nel deserto), e di Giovanni Battista con l'Agnello di Dio; delle decorazioni di derivazione celtica ispirate all'Albero della Vita; delle iscrizioni runiche su tre lati, che tuttavia non ci conservano alcun testo poetico.

Caratteristica unica fra gli oggetti descritti, la presenza di una meridiana ("by far earliest English sundial to survive" Ó Carragáin 2007), che conferma la funzione originaria di *preaching cross* situata all'ester-

no della vicina chiesa e fornisce argomenti alla teoria della natura 'performativa' di queste croci secondo quanto sostenuto da Ó Carragáin.

9.2.6. La croce di Bruxelles

La croce di Bruxelles è un reliquiario e croce processionale risalente alla fine del X – inizi dell'XI secolo; prodotta sul suolo inglese, come testimoniano le iscrizioni in inglese antico, dalla metà del XVII secolo circa è conservata nella Cattedrale dei SS. Michele e Gudula (Bruxelles, Belgio), dove è esposta ancora oggi. La croce è in legno, di dimensioni ben diverse rispetto ai manufatti di Ruthwell e Bewcastle (46.5 x 28 cm), e la parte frontale originariamente era rivestita da una lamina d'oro, arricchita da gemme: questo prezioso ornamento purtroppo è andato perso alla fine del XVIII secolo, si ritiene ad opera di soldati francesi di stanza a Bruxelles nel 1793. Si sono fortunatamente salvati i rivestimenti in argento del retro e dei lati: sulla lamina posteriore sono incisi i simboli dei quattro evangelisti alle estremità delle braccia della croce, mentre al centro si trova l'immagine dell'*Agnus Dei*; ai lati di quest'ultimo, sul braccio orizzontale, si legge chiaramente DRAHMAL – ME WORHTE in caratteri incisi dall'autore del manufatto. Sulla lamina che riveste i lati, spesso all'incirca un centimetro, sono presenti due iscrizioni:

- due versi corrispondenti ai vv. 44 e 48 del Sogno della Croce;
- una dedica da parte dei committenti, i fratelli Æthlmær e Athelwold, al fratello scomparso, Ælfric, che consacra la croce a Cristo in maniera da ottenere la Sua intercessione per la salvezza della sua anima.

Secondo la tradizione, la croce di Bruxelles conteneva un frammento della Vera Croce.

9.2.7. La *cultural matrix* anglosassone

Testi e oggetti sopra citati non solo sono tra i monumenti del periodo anglosassone più conosciuti e studiati, ma sono anche fortemente interconnessi l'uno con l'altro (cfr. § 1 e *infra*), tanto da formare quella che possiamo chiamare "matrice culturale" anglosassone.

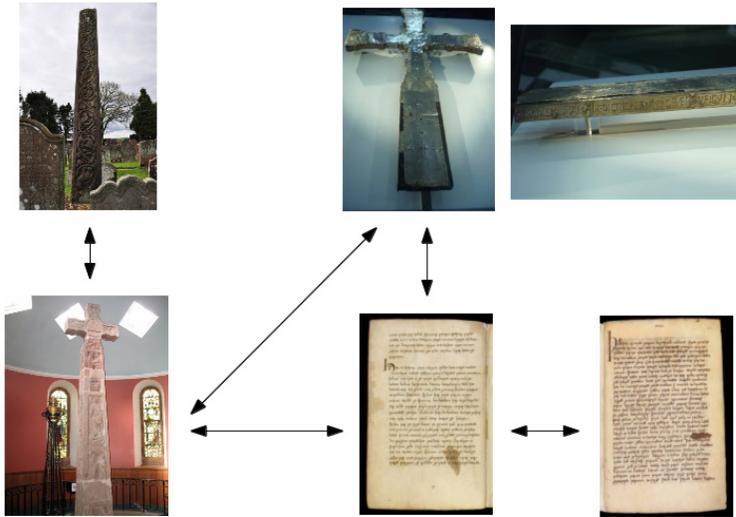


Fig. 9.1. La "cultural matrix" anglosassone: interrelazioni fra le croci e i componenti del Vercelli Book.

Le edizioni a stampa dei singoli testi poetici sono valide, ma ormai risalgono alla seconda metà del secolo scorso: il testo critico più recente del *Sogno della Croce* e dell'iscrizione runica di Ruthwell è Swanton 1970 (rist. 1996), mentre dell'*Elena* abbiamo l'edizione critica curata da P.O.E. Gradon nel 1958 (ed. riv. 1977). Le riproduzioni fotografiche esistenti della croce di Ruthwell sono in genere poco soddisfacenti: l'attuale collocazione della croce in un 'pozzo' caratterizzato da un diametro ridotto (da la forma irregolare si arriva a una distanza minima della croce dalla parete di circa 75 cm) e poco distante dalle pareti della chiesa (non più di 90 cm) rende molto complicata ogni ripresa di tipo analogico, soprattutto dal lato nord; si noti che anche la croce di Bruxelles soffre di problemi simili, in quanto uno dei lati è costituito da una superficie altamente riflettente, e l'iscrizione ha un profilo tale da impedire una buona resa con strumenti di tipo tradizionale; cfr. Ó Carragáin 2005, p. 340.

Gli sviluppi più recenti della tecnologia di acquisizione fotografica (ma non solo, cfr. *infra*) ci hanno convinto che questo fosse il momento ideale per effettuare una nuova edizione di questi oggetti. I problemi tecnici che hanno finora impedito agli studiosi di ottenere riproduzioni fotografiche soddisfacenti sono stati in gran parte risolti: scanner e

fotocamere moderne sono in grado di eseguire una ripresa della croce di Ruthwell in maniera ininterrotta da tutti i lati; nuove tecniche fotografiche permettono di ottenere immagini ad alta risoluzione, prive di riflessi, mappabili su modelli 3D. Grazie a queste tecniche ci aspettiamo di recuperare dettagli altrimenti invisibili a occhio nudo: nel corso di riprese sperimentali della croce di Bruxelles, ad esempio, O'Donnell ha potuto recuperare tracce di quelle che hanno l'aspetto e la forma di incastonature per gemme nel lato perduto di questo manufatto.

Questi progressi tecnologici ci permettono di fare molto di più che non semplicemente migliorare la qualità delle riproduzioni di tali oggetti: ci permettono anche di presentarli in modo innovativo. Le croci monumentali come quella di Ruthwell, in particolare, sono *ipso facto* oggetti multimediali che derivano il loro significato dalla sofisticata interazione fra testo, immagini e contesto. L'iscrizione in versi sulla Ruthwell Cross suggerisce uno sforzo deliberato di adattamento al suo contesto attuale: nei suoi studi più recenti Ó Carragáin sostiene che i parametri spaziali, come la sequenza e la direzione in cui il pubblico di fedeli avrebbe letto le iscrizioni e le scene sui manufatti di Ruthwell e Bewcastle, contribuiscono al significato complessivo di questi oggetti (Ó Carragáin 2005). Facendo uso di un sistema di modellazione 3D gli utenti possono verificare queste teorie, o sperimentare con ricostruzioni alternative riguardo l'aspetto e la funzione originaria di questi monumenti.

Le nuove tecnologie, inoltre, ci consentono di affrontare la ricerca da una nuova prospettiva. Come affermato recentemente da Ó Carragáin, questi capolavori artistici appartengono a una più ampia tradizione relativa alla visione della Croce: se durante il periodo anglosassone qualcuno avesse avuto la fortuna di vedere tutti gli oggetti trattati nel progetto (un'eventualità alquanto improbabile, data la loro distribuzione spaziale e temporale), li avrebbe interpretati sia come opere artistiche individuali, sia come componenti di una rete di tradizioni culturali estesa su piani diversi e che collega oggetti esistenti al di fuori di questo piccolo corpus. La rappresentazione di connessioni intuitive e non strutturate come quelle descritte poco sopra è, naturalmente, la caratteristica fondamentale del funzionamento di Internet e del Web ai giorni nostri. Sulla base di questa cruciale premessa intendiamo costruire un'edizione digitale che permetta agli

utenti di avvicinarsi a queste opere secondo le modalità descritte sopra: considerandole come oggetti individuali dotati di caratteristiche proprie, o seguendo i collegamenti (in questo caso dei link ipermediali) che portano da un oggetto all'altro. In ultima analisi, è questo tipo di innovazione che rende necessarie nuove edizioni: non tanto perché siano datate o inaffidabili quelle esistenti, ma perché il loro angolo visuale è fortemente ristretto al singolo testo o, nel caso del *Sogno della Croce*, al rapporto fra il testo del Vercelli Book e quello della croce di Ruthwell.

Un'edizione digitale, infine, presenta indiscussi vantaggi 'materiali' rispetto a un'edizione a stampa, basti pensare a quanto affermato già da Swanton nella prefazione alla ristampa del 1978 e successive del *Sogno della Croce*: "The high cost of resetting unfortunately precludes the accommodation of valuable points made to me in person and in the continually growing body of critical literature." (Swanton 1978); oggi non possiamo fare altro se non interrogarci su quelli che potrebbero essere i *valuable points* citati da Swanton, ma un'edizione digitale non presenta costi di revisione o aggiornamento, essendo dinamica per definizione è infatti possibile aggiungere correzioni, aggiornamenti ed espansioni al materiale originario *ad libitum*. Grazie alla creazione di una infrastruttura espandibile e alle tecnologie di tipo collaborativo più recenti, pertanto, il nostro progetto, come la rete culturale che rappresenta, può essere ampliato man mano che le nostre conoscenze crescono.

9.3. Acquisizione dei dati 2D/3D

La collaborazione con il progetto Vercelli Book Digitale¹² ha reso disponibile per i ricercatori di Visionary Cross un buon set di foto digitali del *Codex Vercellensis*. Nella prima metà di marzo 2013, inoltre, un team di ricercatori del progetto Lazarus¹³ si è recato a Vercelli per sperimentare le più recenti tecniche di restauro digitale su alcuni

¹² Si veda il sito <http://vbd.humnet.unipi.it/>. Una versione preliminare dell'edizione, che include il testo del *Sogno della Croce*, è adesso disponibile all'indirizzo <http://vbd.humnet.unipi.it/beta/>.

¹³ Pagina Home del progetto disponibile all'URL <http://www.honors.olemiss.edu/lazarus-project/>.

manoscritti dell'Archivio Capitolare: sono state effettuate delle scansioni multispettrali del Vercelli Book che potrebbero rivelare nuovi dettagli, al momento invisibili a occhio nudo, del manoscritto. Sia l'*Elena*, sia il *Sogno della Croce*, dunque, sono disponibili in forma di immagini digitali ad alta risoluzione pronte per essere inserite all'interno dell'edizione, corredate da versioni migliorate, grazie al restauro digitale, di specifici dettagli.

Un primo set di immagini della croce di Bruxelles è a disposizione dei ricercatori del progetto grazie a una scansione sperimentale effettuata da O'Donnell nel 2005: questo manufatto presenta il problema specifico di una superficie riflettente e di aver subito danni in seguito al saccheggio del 1793 (cfr. § 8.2.6). Per quanto non siano da considerarsi definitive le foto che abbiamo mostrato come, facendo uso di fotocamere digitali e di nuove tecniche di elaborazione delle immagini digitali, sia possibile ottenere fotografie ad alta risoluzione, prive di riflessi, in grado di rivelare dettagli preziosi come le tracce residue della forma e della posizione delle gemme originariamente presenti sulla lamina frontale.

I finanziamenti ottenuti nel corso del 2011 hanno permesso quella che, a tutt'oggi, è l'acquisizione dati più significativa per l'avanzamento del progetto, ovvero la scansione 2D e 3D della croce di Ruthwell. Pianificata con cura, la missione si è svolta nell'aprile 2012 nell'arco di soli cinque giorni. Le riprese fotografiche sono state effettuate dal prof. James Graham dell'Università di Lethbridge, mentre la scansione 3D è opera di Marco Callieri e Matteo Dellepiane, ricercatori dell'ISTI-CNR di Pisa. I tre *principal investigator* del progetto erano presenti per preparare il sito (O'Donnell) e successivamente dirigere le scansioni indicando, in particolare, i punti da acquisire con la risoluzione massima dello scanner 3D¹⁴ (Karkov, O'Donnell, Rosselli Del Turco).

Nell'attesa di poter effettuare scansioni 2D e 3D delle croci restanti (Bewcastle e Bruxelles), è in corso la preparazione di una prima edizione della croce di Ruthwell (cfr. *infra*) e la creazione di una infrastruttura interna al progetto che permetta di condividere dati e documenti (repository SVN, area FTP, wiki) e cominciare a definire le caratteristiche delle edizioni future.

¹⁴ Uno scanner a triangolazione laser Minolta Vivid 910, la risoluzione normale è 0,25mm, mentre la massima, usata per dettagli significativi, è 0,17mm.

9.4. Progettare un'edizione integrata

Metodi e tecniche dell'area ICT applicate agli studi umanistici (in breve, l'Informatica Umanistica) non solo permettono di rinnovare l'impostazione classica tipica di molte discipline umanistiche, e di ottenere risultati difficilmente conseguibili con un approccio tradizionalista, ma hanno anche numerose ricadute positive. Tra queste ultime possiamo citare il grande impulso alla divulgazione e conservazione del patrimonio culturale dell'umanità e, particolarmente rilevante ai fini del progetto Visionary Cross, l'abbattimento di molte barriere: su un piano verticale, cronologico, riducendo enormemente i tempi che normalmente intercorrono fra la raccolta e la valutazione dei dati della ricerca, la loro elaborazione, la produzione e diffusione dei risultati scientifici ottenuti; su un piano orizzontale, sono le barriere interdisciplinari a mostrare segni di cedimento, perché molto spesso i dati raccolti risultano utili a studiosi di aree diverse, e li incoraggiano a lavorare gomito a gomito. Questa convergenza di interessi e ricerche finora tenuti separati da rigidi steccati riguarda molto da vicino il nostro progetto: un ampio livello di multidisciplinarietà era implicito sin dall'inizio, essendo il punto di partenza un testo poetico trasmesso sia da una croce monumentale in pietra, che riporta un frammento in caratteri runici, sia da un manoscritto in inglese antico. Ma la disponibilità di un modello 3D della croce chiama in causa anche quelli che vengono chiamati *museum studies* nel mondo anglosassone, e il più recente settore di studi relativo alla realtà virtuale e agli ambienti immersivi 3D: siamo dunque di fronte a una convergenza di queste discipline con la storia dell'arte e la filologia, senza contare l'interesse che i dati raccolti e la loro successiva elaborazione possono rivestire anche per altre aree (ad esempio gli studi di epigrafia, in particolare runica, le discipline paleografiche, che possono avvalersi delle immagini ad alta risoluzione del Vercelli Book, etc.).

L'integrazione dei diversi approcci, l'impulso ad aggiornare metodi e tecniche delle discipline fondamentali coinvolte nel progetto, la preparazione e gestione di dati e risultati in modo che possano essere utilizzati da ricercatori di altre aree, tuttavia, richiedono uno sforzo di organizzazione del progetto particolarmente impegnativo:

- è necessario definire con cautela, e per certi aspetti *ex novo*, la teoria e l'approccio metodologico in grado di garantire i migliori risultati;

- è assolutamente indispensabile procurarsi e utilizzare gli strumenti tecnici più adeguati per l'integrazione e la navigazione/visualizzazione di materiali così diversi; questo include la progettazione di un'interfaccia utente funzionale a tale scopo;
- sarà altrettanto fondamentale, a un punto più avanzato di evoluzione del progetto, definire e inserire nell'infrastruttura il supporto per l'aspetto 'sociale', collaborativo e distribuito, dell'edizione.

Queste esigenze imprescindibili hanno determinato non solo un approfondimento sul piano metodologico, ma una ristrutturazione dell'intero progetto, in particolare della *roadmap* di pubblicazione dei risultati.

9.4.1. Questioni metodologiche

La convergenza di studi filologici, ambienti virtuali e *museum studies*, il tutto sotto il più vasto ombrello delle iniziative volte alla salvaguardia e valorizzazione del patrimonio culturale, è un fenomeno assai recente, che tuttavia sul piano teorico trova una naturale affinità con la corrente di studi testuali nota come *new philology* (Cerquiglini 1989; Nichols 1990): non solo l'interesse è fortemente incentrato sui singoli oggetti, piuttosto che su una ricostruzione considerata anacronistica e arbitraria come quella garantita dal metodo lachmanniano, ma un'edizione digitale, grazie alle sue doti di flessibilità, dinamicità ed espandibilità, sembra anche il mezzo migliore per rendere conto della mobilità del testo nelle varie fasi di produzione e trasmissione; in breve, l'edizione elettronica sembra essere il mezzo ottimale per rendere conto della *mouvance* testuale teorizzata da Zumthor (Zumthor 1972).

Come mostra la frequenza dei termini 'testo' e 'testuale' nel paragrafo precedente, tali considerazioni sembrerebbero applicabili soltanto a oggetti di tipo testuale, come *Il sogno della Croce*: in che modo può essere creata un'edizione di un manufatto come la croce di Ruthwell? cosa significa 'edizione critica' in questo caso specifico? e quale può essere il metodo migliore per visualizzare la fitta rete di interrelazioni che caratterizza le due croci monumentali, la croce di Bruxelles e i testi poetici del Vercelli Book?

Il punto di partenza, per un filologo, è sicuramente il dato testuale,

ed è fin troppo facile notare come la croce di Ruthwell (come quelle di Bewcastle e Bruxelles) sia un supporto per la trasmissione di un testo; non solo, questo testo rientra in un quadro di elaborazione e trasmissione nel corso dei secoli complesso ed affascinante, che si presta sia a una *constitutio textus* di tipo lachmanniano, sia a un'edizione separata di ciascun oggetto ispirata alla *new philology*.

Possiamo tuttavia attribuire alla filologia una funzione più generale, non limitata alla critica testuale: "The fundamental obligation of philology, of the humanities, is the preservation of cultural memory. It is an obligation that has been made both more difficult and more imperative in a world of just-in-time globalized cultural ex-change." (McGann 2012). Quello che risulta fondamentale è il criterio prettamente filologico di verifica: le nostre ipotesi interpretative saranno verificabili (o, per citare Popper, falsificabili) grazie agli strumenti messi a disposizione dell'utente, esattamente come l'apparato critico di un'edizione a stampa mette il lettore in grado di verificare la correttezza della ricostruzione filologica del testo. Quindi il confronto fra il prima/dopo di un'immagine sottoposta a restauro digitale, la possibilità di misurare con precisione le dimensioni delle rune, la ricostruzione delle sezioni danneggiate o della loro posizione originale, e la loro successiva analisi, sono da considerarsi non solo il risultato delle ricerche dei ricercatori del progetto, ma anche un punto di partenza per chi intenda usufruirne per scopi di ricerca scientifica.

9.4.2. Strumenti e fasi del progetto

La versione 'finale' dell'edizione, cui si è accennato più volte nei paragrafi precedenti, presenterà caratteristiche fortemente innovative: volendo essere coerenti con il nome del progetto potremmo dire di avere un 'visione' abbastanza chiara della struttura generale e di quello che verrà offerto all'utente, mentre non è altrettanto definito il percorso e le tecnologie che porteranno al risultato finale. I fattori in gioco sono molti, ed è molto probabile che si debba fare i conti con le limitazioni del software attuale, sarà inevitabile arrivare a compromessi. Un certo grado di incertezza in merito ad alcuni aspetti strettamente legati all'evoluzione della tecnologia è del tutto naturale, come afferma Joris van Zundert "the uncertain and volatile nature of innovation determines that it is hard to establish the forms and requirements of any underlying technology or infrastructure." (van Zundert 2012).

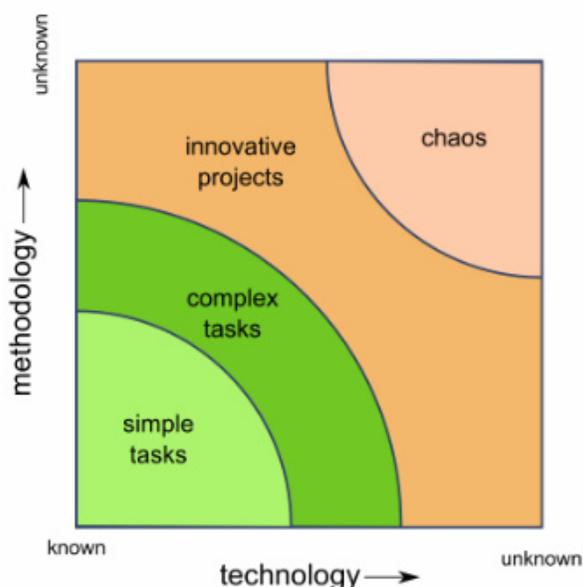


Fig. 9.2. I rischi per i progetti innovativi secondo Joris van Zundert (van Zundert 2012)

Per ridurre al minimo il rischio di investire su soluzioni che successivamente si mostrano inadeguate o solo parzialmente adeguate ai nostri fini, con conseguente spreco di tempo e risorse, abbiamo deciso di adottare una strategia basata su singoli 'passi' del progetto, ciascuno dei quali basato sul ciclo progettazione → creazione di un prototipo → test e raccolta di feedback → implementazione, per coordinare in un secondo momento i vari 'prodotti' ottenuti in sequenza in modo da comporre il disegno più grande.

Il primo passo riguarda la creazione di un'edizione della croce di Ruthwell rivolta in particolare a studiosi e insegnanti di storia dell'arte. Questa primissima edizione sarà basata su un modello a risoluzione medio-bassa e verrà distribuito sul Web grazie a un software sviluppato appositamente, al momento in fase avanzata di sviluppo; seguirà una versione più adatta a fini di ricerca grazie all'inserimento di funzionalità e strumenti specifici (modalità di rotazione e inclinazione più avanzate, misura distanze, etc.).

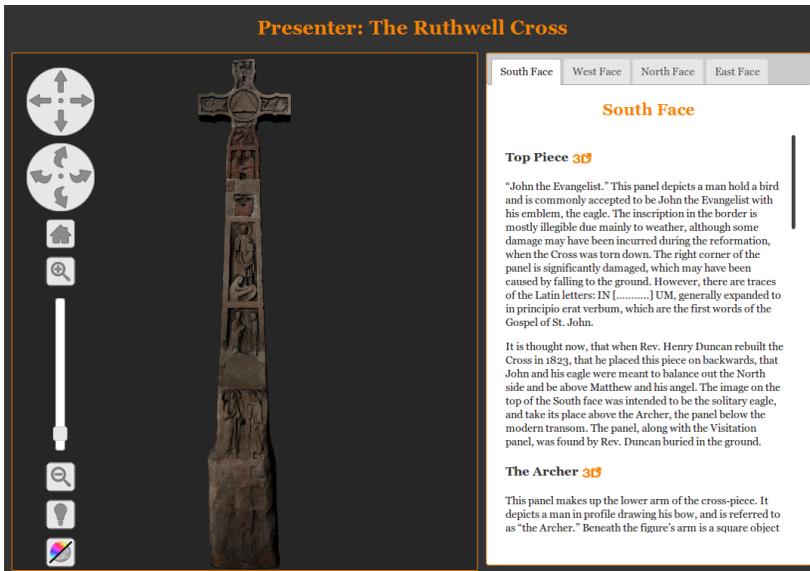


Fig. 9.3. La versione attuale del Presenter 3D

Il passo successivo vedrà l'inserimento dello modello 3D della croce di Ruthwell in un ambiente virtuale basato su tre livelli: la chiesa e il camposanto che la circonda; l'ambiente interno della chiesa; la croce stessa. Quest'ultimo livello si baserà sul software descritto sopra, mentre per i due livelli precedenti stiamo considerando Google Earth (ambiente esterno) e un motore di rendering 3D quale Unreal 3D (ambiente interno). Il motivo per cui stiamo valutando ben tre ambienti diversi dipende dalla ricerca della massima economia ed efficienza, per ciascun ambiente viene scelto il software che offre un risultato efficace per la risoluzione richiesta; si tratta in ogni caso di software open source, in modo da poterli integrare con il software prodotto all'interno del progetto, che sarà tutto distribuito con licenza *open source*.

9.4.3. Obiettivo finale: un'edizione multimediale, collaborativa, distribuita

Se il testo della croce di Ruthwell perde gran parte del suo significato se analizzato in maniera avulsa dal suo contesto (cfr. supra), lo stesso si può dire della croce stessa: il rapporto con la comunità locale era molto forte in epoca medievale, quando le *preaching cross* come quelle di Ruthwell e Bewcastle erano al centro della comunità di

credenti, ma anche in epoca più recente, il che è alquanto sorprendente considerando il conflitto interreligioso che aveva portato allo smantellamento della croce nel XVII secolo. Tuttavia il ritrovamento dei frammenti rappresenta una versione locale della *inventio crucis*, e il legame affettivo con la comunità locale persiste fino ai giorni nostri, come abbiamo potuto testimoniare di persona¹⁵. La parrocchia di Ruthwell, inoltre, è parte di una comunità estesa che comprende anche i discendenti di numerose famiglie emigrate negli Stati Uniti nel corso del XIX secolo: buona parte dei turisti che visitano Ruthwell vengono dal Nord America e tra questi vi sono persone che cercano di riallacciare i fili di legami lontani, in alcuni casi anche visitando il camposanto che circonda la chiesa di Ruthwell in modo da individuare le tombe di lontani parenti.

Questo tipo di ricerca può trovare una valida alternativa in un sito partecipativo, la realizzazione di un ambiente virtuale per esplorare la chiesa di Ruthwell e l'ambiente circostante ha come obiettivo l'interazione (possibilità di annotare il contenuto, commentarlo, condividerlo con altri utenti) con gli oggetti inseriti nella ricostruzione, ad esempio le lapidi delle tombe, spesso caratterizzate da epitaffi contenenti testi di grande fascino oltre che di notevole interesse per quanto riguarda la storia locale. Questo tipo di interazione, che può comprendere buona parte degli oggetti rilevanti per la storia della parrocchia di Ruthwell e che vede la croce quale punto focale, può risultare molto efficace nel creare un ponte fra comunità locale e comunità estesa.

Un'edizione collaborativa, inoltre, presenterebbe molti benefici per la comunità accademica: se oggi possiamo soltanto rimpiangere la perdita delle osservazioni e note riguardo la croce di Ruthwell raccolte da Swanton, la possibilità di annotare testi e modelli dell'edizione, di entrare in contatto con altri studiosi per uno scambio di idee, di verificare le proprie ipotesi manipolando direttamente gli oggetti dell'edizione, sarà di grandissimo aiuto per lo studio e la ricerca.

Come ultima nota, il *framework* dell'edizione sarà completamente aperto: sia dal punto di vista dell'espandibilità, in quanto è prevista sin

¹⁵ Tale sensazione si basa non solo sull'affluenza di visitatori incuriositi dalle nostre operazioni di acquisizione 2D/3D, ma anche sul fatto che ben due giornali locali hanno inviato dei corrispondenti locali ad intervistarci e hanno dedicato ampio spazio al nostro lavoro.

d'ora la possibilità di aggiungere nuovi testi e manufatti all'edizione; sia perché non è indispensabile che tutti gli oggetti siano raccolti sul medesimo server: il software sarà disponibile con licenza *open source*, per cui è ipotizzabile che venga usato per altri testi e modelli 3D, e nel caso vi sia un'affinità con l'edizione Visionary Cross sarà possibile inserire dei collegamenti in modo da far dialogare a distanza oggetti che si trovano su siti web diversi (edizione distribuita); sia perché è previsto l'accesso diretto agli oggetti nella modalità cosiddetta *unnegotiated access*, ovvero senza che sia necessario richiedere preventivamente il permesso agli autori dell'edizione (ovviamente rispettando i termini della licenza che sarà utilizzata per questo scopo), in maniera tale da incorporarli nella propria ricerca o edizione.

9.5. Conclusioni

Dopo un lungo periodo di discussione e pianificazione, la scansione della croce di Ruthwell ha permesso al progetto di compiere un significativo passo in avanti e di cominciare il lungo percorso verso la creazione di un'edizione globale di tutti gli oggetti descritti nei paragrafi precedenti. La decisione di procedere attraverso una serie di fasi e passi intermedi è la strategia che più di ogni altra ci permetterà di limitare al minimo i rischi connessi con la sperimentazione di tecnologie e soluzioni tecniche innovative; le "edizioni intermedie", viceversa, promettono di essere risultati validi e interessanti di per sé nell'ambito degli obiettivi prefissi: la prima di queste, l'edizione della croce di Ruthwell per studiosi e insegnanti di storia dell'arte anglosassone, è in una fase di sviluppo molto avanzata ed è già stata presentata in via preliminare al convegno di studi medievali di Leeds (International Medieval Congress 2013, Leeds 1-4 luglio 2013).

9.6. Bibliografia

9.6.1. Facsimile ed edizioni

- DICKINS B., ROSS, A.S.C. (1963). *The Dream of the Rood*. Methuen's old English library. London: Methuen.
- GRADON P.O.E. (1977). *Cynewulf's 'Elene'*. Exeter Medieval English Texts. Ristampa con revisione dell'edizione Methuen del 1958. Exeter: University of Exeter Press.
- SISAM, C. (1976). *The Vercelli Book: a late tenth-century manuscript containing prose and verse*. Copenhagen: Rosenkilde and Bagger.
- SWANTON M. (1970). *The Dream of the Rood*. Old and Middle English texts. Manchester - New York: Manchester University Press; Barnes & Noble. (rist. 1978 Exeter Medieval English Texts and Studies, Exeter University Press).

9.6.2. Articoli e monografie

9.6.2.1. Ecdotica, cultura anglosassone e leggenda della vera Croce

- BODDEN M.-C. (1987). *The Old English Finding of the True Cross*. Cambridge [England] - Wolfeboro [N.H. USA]: D.S. Brewer.
- BROWN G.H., JOLLY K. L., KARKOV C.E., KEEFER S.L. (2008). *Cross and culture in Anglo-Saxon England: Studies in honor of George Hardin Brown*. Morgantown: West Virginia University Press.
- CERQUIGLINI B. (1989). *Éloge de la variante*. Paris: Cerf.
- COLE B. (1968). *A Reconstruction of Masolino's True Cross Cycle in Santo Stefano, Empoli*. «Mitteilungen des Kunsthistorischen Institutes in Florenz», 13(3/4), pp. 289-300.
- DARLING G.J. (2010). *Cross Legends and Crossings: Links Among Anglo-Saxon, Medieval Irish, and Late Classical Texts*. «Eolas: The Journal of the American Society of Irish Medieval Studies», vol. 4, pp. 98-111.
- DRIJVERS H. (1997). *The finding of the True Cross: the Judas Kyriakos legend in Syriac: introduction, text, and translation*. Lovanii: In aedibus Peeters.
- FÖRSTER M. (1913). *Der Vercelli Codex CXVII: nebst Abdruck einiger Homilien der Handschrift*. Halle: M. Niemeyer.

- GALLOWAY A. (1994). *Dream-Theory in The Dream of the Rood and The Wanderer*. «The Review of English Studies», New Series, vol. 45 (180), pp. 475-485.
- HALSALL M. (1969). *Vercelli and the Vercelli Book*. «PMLA», vol. 84 (6), pp. 1545-1550.
- HARBUS A. (1994). *Text as revelation: Constantine's dream in Elene*. «Neophilologus», vol. 78 (4), pp. 645-653.
- HOWLETT D.R. (1974). *Two Panels on the Ruthwell Cross*. «Journal of the Warburg and Courtauld Institutes», vol 37, pp. 333-336.
- KARKOV C.E. (1991). *Studies in insular art and archaeology*. Oxford (Ohio): American Early Medieval Studies and the Miami University School of Fine Arts.
- KARKOV C.E. (2003a). *Anglo-Saxon styles*. Albany: State University of New York Press.
- KARKOV C.E. (2003b). *Theorizing Anglo-Saxon stone sculpture. Morgantown [W. Va.]*: West Virginia University Press.
- KARKOV C.E. (2006). *The place of the cross in Anglo-Saxon England*. Woodbridge: Boydell Press.
- KARKOV C.E. (2011). *The art of Anglo-Saxon England*. Woodbridge, Suffolk: Boydell Press.
- LUISELLI FADDA A.M. (1994). *Tradizioni manoscritte e critica del testo nel Medioevo germanico*. Bari: Laterza.
- LUISELLI FADDA A.M. (1998). *Osservazioni sulla genesi e sulla stuttura del Dream of The Rood*. In *Vercelli Tra Oriente ed Occidente, tra tarda Antichità e Medioevo*, a c. di V. Corazza, pp. 101-115. Vercelli: Edizioni Dell'Orso.
- LUISELLI FADDA A.M. (2000). *La Croce nella tradizione poetica anglosassone (secc. VIII-X)*. «Romanobarbarica», febbraio 2000, pp. 333-59.
- NESTLE, E. 1889. *De Sancta Cruce. Ein Beitrag Zur Christlichen Legendengeschichte Von E. Nestle*.
- NICHOLS, S.G. (ed.). (1990). *The New Philology*. Special issue of «Speculum» 65, n. 1, pp. 1-108.
- Ó CARRAGÁIN É. (1975). *The Vercelli Book as an ascetic florilegium*. Belfast: Queen's University.

- Ó CARRAGÁIN É. (1978). *Liturgical Innovations Associated with Pope Sergius and the Iconography of the Ruthwell and Bewcastle Crosses*. In *Bede and Anglo-Saxon England papers in honour of the 1300th anniversary of the birth of Bede, given at Cornell University in 1973 and 1974, British archaeological reports* (Vol. Series Volume: 46, pp. 131-147). London: British Archaeological Reports.
- Ó CARRAGÁIN É. (1987-88). *The Ruthwell Crucifixion Poem in Its Iconographic and Liturgical Contexts*. «Peritia» vol. 6-7, pp. 1-71.
- Ó CARRAGÁIN É. (1998). *Rome, Ruthwell, Vercelli: "The Dream of the Rood" and the Italian Connection*. In *Vercelli Tra Oriente ed Occidente, tra tarda Antichità e Medioevo*, a c. di V. Corazza, pp. 59-100. Vercelli: Edizioni Dell'Orso.
- Ó CARRAGÁIN É. (2005). *Ritual and the rood: liturgical images and the Old English poems of the Dream of the rood tradition*. Toronto: University of Toronto Press.
- Ó CARRAGÁIN É. (2007). *Christian Inculturation in Eighth-Century Northumbria: The Bewcastle and Ruthwell Crosses*. «Colloquium Magazine», vol. 4, Yale Institute of Sacred Music. http://www.yale.edu/ism/colloq_journal/vol4/carragain1.html.
- REUTER T., KEEFER S.L., JOLLY K.L., KARKOV C.E. (2010). *Cross and cruciform in the Anglo-Saxon world: Studies to honor the memory of Timothy Reuter*. Morgantown: West Virginia University Press.
- SCRAGG D.G. (1973). *The compilation of the Vercelli Book*. «Anglo-Saxon England», vol. 2, pp. 189-207 (rist. in Richards M. P. (1994) *Anglo-Saxon Manuscripts: Basic Readings*. New York: Garland).
- SCRAGG D. G. (1992). *The Vercelli Homilies and Related Texts*. Early English Text Society (Series), no. 300. Oxford: Oxford University Press.
- THIEME A.L.J. (1998). *Gift Giving as a Vital Element of Salvation in 'The Dream of the Rood'*. «South Atlantic Review», vol. 63 (2), pp. 108-123.
- ZACHER S. (2009). *New Readings in the Vercelli Book*. Toronto Anglo-Saxon series Vol. 4. Toronto-Buffalo, NY: University of Toronto Press.
- ZUMTHOR P. (1972). *Essai de poésie médiévale*. Collection Poétique, Paris: Seuil.

9.6.2.2. Digital Humanities

- CLEMENT T. (2011). *Knowledge Representation and Digital Scholarly Editions in Theory and Practice*. «Journal of the Text Encoding Initiative» [Online], Issue 1 June 2011. <http://jtei.revues.org/203>.
- FINNERAN R.J. (1996). *The literary text in the digital age*. Ann Arbor: University of Michigan Press.
- FIORMONTE D., USHER J. (a cura di) (2001). *New Media and the Humanities: Research and Applications*. Oxford: Humanities Computing Unit.
- KARLSSON L., MALM L. (2004). *Revolution or remediation? A study of electronic scholarly editions on the web*. «HumanIT» vol. 7.1, pp. 1-46. <http://www.hb.se/bhs/ith/1-7/klm.pdf>.
- KIERNAN K.S. (2011a). Edition Production & Presentation Technology (EPPT) <http://beowulf.engl.uky.edu/eppt/>.
- KIERNAN K.S. (2011). *Electronic Beowulf* [CD-ROM]. Third edition. London: British Library.
- MCGANN J. (2004). *Marking Texts of Many Dimensions*. In *Companion to Digital Humanities* (Blackwell Companions to Literature and Culture), a c. di R. Siemens, J. Unsworth, e S. Schreibman, pp. 198-217. Oxford: Blackwell Publishing Profession.
- MCGANN J. (2012). *Memory Now*. «4Humanities». <http://4humanities.org/2012/08/jerome-j-mcgann-memory-now-2/>.
- O'DONNELL D.P. (2004). *The Doomsday Machine, or, 'If you build it, will they still come ten years from now?': What Medievalists working in digital media can do to ensure the longevity of their research*. «Heroic Age» vol. 7. <http://www.mun.ca/mst/heroicage/issues/7/ecolumn.html>.
- O'DONNELL D.P. (2005a). *Cædmon's Hymn: A multimedia study, archive and edition*. Society for early English and Norse electronic texts A.7. Cambridge and Rochester: D.S. Brewer in association with SEENET and the Medieval Academy.
- O'DONNELL D.P. (2005b). *O Captain! My Captain! Using Technology to Guide Readers Through an Electronic Edition*. «Heroic Age 8». <http://www.mun.ca/mst/heroicage/issues/8/em.html>.
- O'DONNELL D.P. (2007). *Disciplinary impact and technological obsolescence in digital medieval studies*. In *A companion to digital literary studies*, a c. di S. Schreibman e R. Siemens. Oxford: Blackwell, pp. 65-81.

- ORE E.S. (2004). *Monkey Business—or What is an Edition?* «Literary and Linguist Computing», vol. 19, pp. 35-4.
- ROBINSON P. (2002). *What is an Electronic Critical edition?* «Variants», vol. 1, pp. 43-62.
- ROBINSON P. (2004). *Where we are with electronic scholarly editions, and where we want to be.* «Jahrbuch für Computerphilologie Online» 1.1. <http://computerphilologie.uni-muenchen.de/jg03/robinson.html>.
- ROBINSON P. (2005). *Current issues in making digital editions of medieval texts—or, do electronic scholarly editions have a future?* «Digital Medievalist» vol. 1.1. <http://www.digitalmedievalist.org/article.cfm?RecID=6>.
- SHILLINGSBURG P.L. (1986). *Scholarly editing in the computer age: Theory and practice.* Athens: University of Georgia Press.
- SIEMENS R., ET AL. (2012). *Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media.* «Literary and Linguistic Computing» vol. 27, n. 4, pp. 445–461.
- VANHOUTTE E. (1999). *Where is the Editor? Resistance in the creation of an electronic critical edition.* «Human IT», vol. 3(1), pp. 197-214. <http://etjanst.hb.se/bhs/ith/1-99/ev.htm>.
- VAN ZUNDERT J. (2012). *If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities.* «Historical Social Research» vol. 37.3 Special Issue a c. di M. Thaller: *Controversies around the Digital Humanities.*

9.6.3. Siti Web

- Visionary Cross Project. <http://visionarycross.org/>.
- Vercelli Book Digitale. <http://vbd.humnet.unipi.it/>. Versione preliminare dell'edizione: <http://vbd.humnet.unipi.it/beta/>
- Digital Medievalist. An international web-based Community of Practice for medievalists working with digital media. <http://www.digitalmedievalist.org/>.
- MENOTA (MEdieval and NORse Text Archive), publishers of the Menota handbook: Guidelines for the encoding of medieval Nordic primary sources. <http://www.menota.org/>.
- MUFI (Medieval Unicode Font Initiative), an organisation dedicated

to the development of solutions to character encoding issues in the representation of characters in medieval manuscripts. <http://gandalf.aksis.uib.no/mufi/>.

TEI (Text Encoding Initiative). An international and interdisciplinary standard that enables libraries, museums, publishers, and individual scholars to represent a variety of literary and linguistic texts for online research, teaching, and preservation. <http://www.tei-c.org/>.

10. Many witnesses, many layers: the digital scholarly edition of the *Iudicium coci et pistoris* (Anth. Lat. 199 Riese)

*Paolo Monella*¹

Abstract. This article will describe the rationale of the digital scholarly edition on which I am currently working (the *Iudicium coci et pistoris* by Vespa, *Anth. Lat.* 199 Riese), an ancient Latin text with a multi-testimonial textual tradition. My edition aims to provide a proof-of-concept application of the ideas of Tito Orlandi and Raul Mordenti while using a customised XML/TEI markup.

In my edition, each witness will be encoded at three layers:

1. the graphic layer (graphemes and other graphic signs);
2. the alphabetic layer (alphabetic letters);
3. the linguistic layer (inflected words).

The main proposed innovations are as follows:

1. the distinction of different textual layers within each witness;
2. the declaration of 'tables of signs' for the graphical and the alphabetical layers of each witness;
3. each layer of a witness will be collated with the corresponding layers of other witnesses.

Keywords: textual layers; encoding levels; table of signs; collation; characters; graphemes; alphabemes; manuscripts; XML/TEI.

¹ Centro Linceo Interdisciplinare "B. Segre", Accademia dei Lincei, Rome, Italy.
e-mail: paolo.monella@gmx.net.

10.1. The conceptual framework

10.1.1. The text edited

I am currently working on a scholarly digital edition of the *Iudicium coci et pistoris iudice Vulcano* by Vespa, a Latin mock court debate in verse between a cook and a baker, written in the Roman late imperial age and included by Riese in his *Anthologia Latina* as poem number 199².

The edition will contain a number of experimental features. While its realisation is still in progress, the aim of this article is to reveal and discuss the rationale of these innovations, as well as the open issues which arise from them³.

10.1.2. A proof of concept

This edition aims to be a proof of concept. It is my aim to ascertain whether the sophisticated theoretical and methodological reflections of Tito Orlandi and Raul Mordenti on digital scholarly editions may be implemented in a prototype through a sustainable work-flow⁴. My experiment addresses two issues central to the current debate in digital philology: manuscript encoding and manuscript collation. The main ideas by Tito Orlandi that I endeavour to apply are⁵:

² Its oldest witnesses are the *Codex Salmasianus* (Parisinus 10318, VII or VIII Century) and the *Codex Thuaneus* (Parisinus 8071, IX/X Century). Modern editions, after Bücheler and Riese 1894-1926, include Pini 1958, Shackleton Bailey 1980, Baumgartner 1981, Shackleton Bailey 1982 (where the poem is number 190), Barry 1987 and Lespect 2005. Omont 1903 is a photographic reproduction of *Codex Salmasianus*, but the codex is now fully digitised and openly accessible in Gallica at <http://gallica.bnf.fr/ark:/12148/btv1b8479004f>.

³ I wish to thank James Pearson-Jadwat for his precious advice, going well beyond a mere linguistic revision of my English.

⁴ See (at least) Orlandi 1999, Orlandi 2010 and his *Edizione digitale sperimentale di Niccolò Machiavelli, De principatibus* (<http://rmcisadu.let.uniroma1.it/~orlandi/principe>, last retrieved 10.03.2013); Mordenti 2001, Mordenti 2012 and his *Edizione Critica Ipertestuale dello Zibaldone Laurenziano (Pluteo XXIX.8) autografo di Giovanni Boccaccio* (<http://rmcisadu.let.uniroma1.it/boccaccio>, last retrieved 10.03.2013).

⁵ See particularly Orlandi 2010.

1. the encoding of each witness's text *at different textual layers*, including the graphical, alphabetic and linguistic ones;
2. the explicit declaration of a *table of signs* for each of the 'lowest' layers (graphical and alphabetic) of each witness;
3. *collation between witnesses by layer*: e.g. the linguistic layer of MS A should be collated with the linguistic layer of MS B, and the alphabetic layer of the former with the alphabetic layer of the latter.

As my aim is to produce a prototype, the edition model came before (and was appropriate to) the choice of text to edit. I chose the *Iudicium coci et pistoris* because

1. It is an ancient Latin work with a multi-testimonial textual tradition – as we, quite remarkably, have no digital scholarly edition of 'canonical' texts of classical antiquity⁶.
2. It has a (small) multi-testimonial handwritten textual transmission – so I can devise mechanisms for collation between witnesses with different writing systems.

10.1.3. Textual layers

In my edition, I have decided to encode the text of each manuscript at three different textual layers:

1. the 'graphical' layer, whose minimal units of encoding are graphemes and paragraphematic signs (like punctuation);
2. the 'alphabetic' layer, whose minimal units are alphabetic letter ('*alphabemes*', from now on⁷);
3. the 'linguistic' layer, whose minimal units are inflected words.

The choice of these three layers among the many others possible (lemmatical, allographic, idiographic, etc.⁸) is arbitrary, and responds to the scientific purposes of a specific edition⁹. Ideally, a digital edition

⁶ I propose an epistemological explanation for this in Monella, forthcoming.

⁷ The term '*alphabeme*' was suggested to me by Raul Mordenti in an email in December 2012.

⁸ I am here referring to the terminology fixed by Peter Stokes for the DigiPal Project (<http://www.digipal.eu/blogs/blog/describing-handwriting-part-iv>, last retrieved 10.03.2013).

⁹ For instance, the '*allographic*' and '*idiographic*' layers are relevant for Mordenti's edition of the *Zibaldone Laurenziano* by Boccaccio (see footnote 2 above).

should be a modular and 'open source' digital object open to integrations by other scholars: another editor should be able, for example, to add an 'allographic' layer to my edition, if he so chooses¹⁰.

10.1.4. Graphemes and alphabemes

It should firstly be pointed out that graphemes and alphabemes (alphabetic letters) are not the same thing¹¹. In the *Glossary of Unicode Terms*¹², a "letter" (here called an 'alphabeme') is defined with no reference to a graphical representation as "An element of an alphabet", while a grapheme is defined as "A minimally distinctive unit of writing in the context of a particular writing system".

In my terminology, both grapheme 'j' and the Morse code / · - - - / represent alphabeme 'j'¹³. Likewise, allographs like 'capital J' and 'lowercase j' represent grapheme 'j', and idiographs like my individual handwritten sign for 'lowercase j' represent allograph 'lowercase j'.

10.1.5. Five Gutenbergian assumptions

Our approach to digital textual encoding tends to be influenced by the standardisation of graphic systems brought about by print in the last centuries, as well as by a number of related assumptions that are simply not valid for ancient manuscripts:

¹⁰ On the concept of an open source critical edition, see Bodard and Garcés 2009.

¹¹ See Emiliano 2011, 154-157, where a clear distinction is drawn between "letter" (my 'alphabeme'), "character" and "glyph".

¹² <http://www.unicode.org/glossary>, last retrieved 10.03.2013.

¹³ An example might clarify my conceptual distinction. Latin 'j' is an *alphabeme* in that it belongs to the modern English *alphabet* (but not to the traditional Italian alphabet that I learnt in my first grade). It also belongs to the Spanish and to the French alphabet, but in each of them it corresponds to a different *phoneme*. An alphabeme is an abstract cultural concept. Latin alphabeme 'a' is different to Greek 'alpha'. If I want to convey the message 'alphabeme j', either as a part of a spelled word or as the name of a specific section of a tax refund form, I can use a range of signs that all *represent* that alphabeme (but obviously *are not* that alphabeme): I can write my idiograph for the corresponding *grapheme* 'j' on a piece of paper, I can (rather unknowingly) enter in a computer the Unicode codes U+006A (small) or U+004A (capital), I can use the corresponding fingerspelling handshape of American Sign Language, I can pronounce the phonetic chain /dʒeɪ/ or, if I am a sophisticated student in search of ingenious ways to cheat on exams, I can use the Morse code / · - - - /.

1. *Standard alphabet*: all witnesses of a text share a standard alphabet (a set of alphabemes).
 - A counterexample might be a digital edition whose witness base includes a medieval manuscript with no distinction between alphabemes 'u' and 'v', and a modern print edition which makes that distinction. Likewise, some Middle English manuscripts include the 'thorn' alphabeme ('þ') and some do not.
2. *Standard graphic system*: all witnesses of a text share a standard graphic system (a set of graphemes, punctuation, capitalisation conventions etc.).
 - On the basis of the Unicode definition of a grapheme mentioned above, I consider handwritten systematic abbreviations (like 'ē' for 'em') to be graphemes. Such abbreviation conventions vary greatly between manuscripts and normally do not exist in modern print editions. Other aspects of graphic systems which were not standardised until the invention of print include punctuation and comparable forms of 'graphic markup'.
3. *Standard spelling*: a specific sequence of alphabemes (e.g. 'w', 'i', 'f' and 'e') can be taken as a standard representation of an inflected word (e.g. the singular of 'wife'). Put simply, there exists a standard spelling for each word.
 - Spelling is notoriously variable at all stages of a language's historical development. Even in modern European languages, spelling was (almost) completely standardised only recently. Also, old variant spellings of English still exist ('color'/'colour') and new ones keep emerging ('night'/'nite'). Latin, to name another Western *interlingua* that at some point was fixed in a 'standard' form, features 'deviant' spellings in archaic inscriptions and in the treatment of diphthongs like 'ae'.
4. *Standard sequentiality*: graphemes in a written text form are in an ordered sequence flowing in one direction only.

- A counter example is provided by the Devanagari Indic script. It normally flows from left to right, but a vowel that phonetically follows a consonant may be written 'before' (to the left of) that consonant¹⁴. A case more familiar to Western scholars is that of Greek iota subscript ('ι'), but many more examples could be taken from the Arabic script conventions for vowels and from the medieval European custom of writing some letters above others.
5. *One grapheme, one alphabeme*: there exists a one-to-one correspondence between alphabemes and graphemes in writing.
- This is not only contradicted by the use of ideographs and logographs in Western scripts (like '&' or the Tyronian note), but also by the existence of alphabetic writing systems that systematically do not write a grapheme for each alphabeme, like Arabic and Hebrew, or that make an extensive use of systematic abbreviations, like ancient and medieval Latin and Greek writing systems before the invention of print. In the latter case, as noted above, I consider the final abbreviation in 'regē' as one grapheme ('ē') corresponding to two alphabemes ('e' and 'm'). I should emphasise further that in such cases the correspondence of *one* grapheme to *many* alphabemes was *systematic*: this is how a medieval scribe learned to write, and he did so even when writing or copying an important or sacred text on a luxury codex. It is also worth noting that assumption no. 5 is probably the reason why contemporary Western literates find it hard to distinguish between the concepts of alphabeme and grapheme.

10.2. TEI issues and solutions

10.2.1. Graphic/alphabetic layers in TEI?

Does TEI already provide mechanisms to distinguish formally between the graphic and alphabetic layers when encoding a text to which the assumptions above do not apply? Let us consider the case of abbreviations. It is possible in principle that a theoretically

¹⁴ See Fiormonte 2012a, 68 (corresponding to page 9 in the PDF file in http://www.cceh.uni-koeln.de/files/Fiormonte_final.pdf) and Perri 2009, 736.

conscious and consistent use of elements like `<abbr>` and `<ex>` might provide means to encode a witness's grapheme/alphabeme distinction in a concise and 'economic' though formally unambiguous way. However, due to the (intentionally) loose definition of those elements in the *Guidelines* and to the different encoding conventions allowed, the current TEI encoding of abbreviations – and above all its practical application in extant editions – does not appear to provide a reliable mechanism for that task¹⁵.

Indeed, the issue is more general: simply, the TEI P5 Guidelines do not postulate any distinction between graphemes and alphabemes. The general concept of 'character' in TEI P5 assumes a triple correspondence one grapheme in the document, one alphabetic letter in the abstract text, one digital code point in the XML file¹⁶.

10.2.2. A Saussurean issue

This brings us to a second, more complex, issue. The TEI Guidelines strongly recommend that in Digital Humanities projects, 'characters' are encoded *according to the Unicode standard*¹⁷. The Unicode system, in turn, is based on the fact that a 'LATIN SMALL LETTER U' is a 'u' (i. e., despite the name 'letter', the same grapheme) in all written documents, from Late Antiquity parchment codices to contemporary websites, and as

¹⁵ Methodological reflections on digital 'diplomatic' and 'normalised' editions like Driscoll 2006 and Pierazzo 2011 are not based on the grapheme/alphabeme distinction.

¹⁶ The two key sections of the TEI P5 Guidelines on the encoding of 'characters' are *vi. Languages and Character Sets* (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CH.html>) and *5. Representation of Non-standard Characters and Glyphs* (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/WD.html>). Section *vi* (#D4-42) makes a useful distinction between "abstract characters" (which seem to coincide with my 'graphemes') and "glyphs" (seemingly 'allographs' in Stoke's and my own terminology, i.e. graphic variants of the same 'abstract character', with no distinctive value). However, I still can neither find any conceptual distinction between grapheme and alphabetic letter (my 'alphabeme') in that section, nor elsewhere in the TEI P5 Guidelines. Compare Orlandi 2010, 8-9 and 48; Emiliano 2011, 154-157.

¹⁷ I am now using 'character' and 'glyph' according to the TEI P5 (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CH.html#D4-42>) and Unicode terminology (<http://unicode.org/glossary/>); both links were last retrieved on 19.03.2013. The 'Unicode-compliance principle' is described in sections *vi. Languages and Character Sets* and *5. Representation of Non-standard Characters and Glyphs* of the TEI Guidelines (mentioned in footnote 5) and discussed in Wittern 2006.

such it should be encoded with the same code point (U+0075).

This 'Unicode-compliance' principle works fairly well for the digitisation of those post-Gutenberg print documents that feature a standardised alphabet, graphic system, spelling, sequentiality and a one-to-one correspondence between grapheme and alphabeme. However, this principle does not suffice *per se* as a general principle for the encoding of medieval manuscripts or other documents that do not follow our 'standard' graphic system. Even worse does this principle serve us when we try to collate (or search through) a set of those documents¹⁸. The reason for this is that a 'u' is *not* always just a 'u': Ferdinand de Saussure taught us that signs have a relational nature *within* a specific semiotic system. This is echoed by the Unicode definition of a grapheme as "A minimally distinctive unit of writing *in the context of a particular writing system*"¹⁹.

Imagine two medieval English manuscripts, A and B: both are copies of the same text, and a philologist is building a digital scholarly edition upon them. The graphic system of MS A has a distinction between grapheme 'u' and grapheme 'v', while MS B does not have that distinction and uses one grapheme ('u') to represent both alphabeme 'u' and alphabeme 'v'²⁰. MS A reads: "The loue of love" ('the hill of love'). MS B reads: "The loue of loue". At the linguistic layer, this is the same text, just encoded with different graphemes.

If we encoded both manuscripts following the TEI Unicode-compliance principle, the result would be as shown in Fig. 10.1:

¹⁸ See Orlandi 2010, 9 and 48.

¹⁹ The Unicode definition of "grapheme" is in the *Glossary of Unicode Terms* (<http://unicode.org/glossary>). See Orlandi 2010, 9 and Emiliano 2011, 157-158: "Now consider a similar proposition: 'A is a grapheme'. It is neither true nor false, it is simply meaningless. Because the grapheme (like the phoneme) is a linguistic relational concept, this proposition can only have truth or falsehood content in relation to a specific language. The definition of 'emic' units, contrary to 'etic' units, is a function of their status in a given symbolic system [...] The grapheme, like the phoneme, is an abstract unit, whose value is defined in terms of the relation between elements of the same type in a system".

²⁰ In their turn, 'u' and 'v' are distinct alphabemes because in the Middle English *linguistic* system there exists at least a pair of words that are distinct by the 'u/v' difference: an example is the pair 'loue' ('hill', in modern English) *vs.* 'love' ('love'), which I am using here as an example.

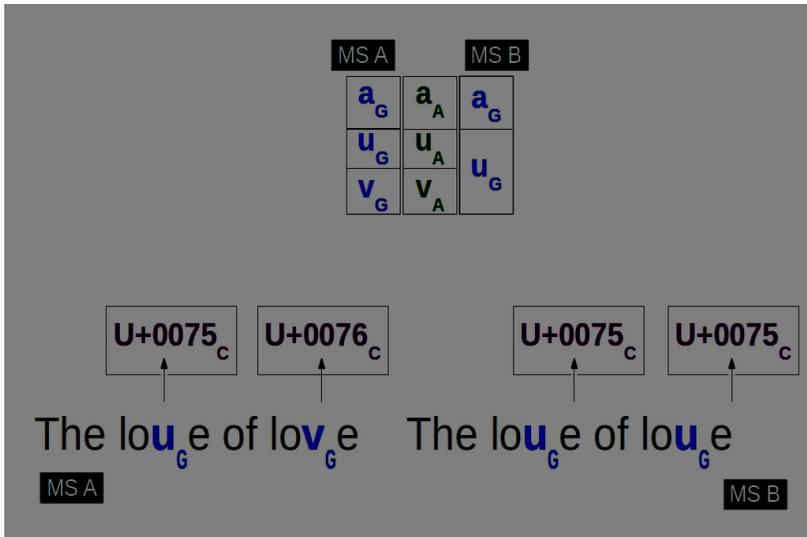


Fig. 10.1. Encoding of graphemes from documents with different graphic systems by use of the TEI 'Unicode-compliance' principle. 'G' stands for grapheme, 'A' for alphabeme and 'C' for code point.

This entails that a piece of software for collation or cross-corpus searching could not avoid being deceived by the apparent identity of the 'u' grapheme in MS A and the 'u' grapheme in MS B (as both are encoded with the Unicode code point U+0075). Similarly, it may appear to it that the 'v' grapheme in MS A is *not the same* as the 'u' grapheme in MS B (as they are encoded with different Unicode code points, U+0075 and U+0076). This is false, as the 'u' grapheme in MS A, defined contrastively by its opposition to the 'v' grapheme, is *not the same* grapheme as the 'u' grapheme in MS B, which has no such opposition and represents both alphabeme 'u' and alphabeme 'v'.

10.2.3. My proposed solutions: the 'musical score' model and tables of signs

As discussed in the previous two paragraphs, if I want to use TEI P5 to encode my manuscripts at different layers with a formal distinction between graphemes and alphabemes, I have two main issues to solve:

1. The layers distinction issue (see 9.2.1. *Graphic/alphabetic layers in TEI?* sopra). My proposed solution is a model of digital scholarly edition resembling a musical score, with three parallel sequences of aligned tokens (graphemes, alphabemes and inflected words). The general model is discussed in 9.3.1. *The abstract data model* sotto. Possible linearisations of this data model will be expanded upon in 9.3.4. *Separate files linearisation model* and 9.3.5. *Menota linearisation model*.
2. The 'Saussure' issue (see 9.2.2. *A Saussurean issue* sopra). My proposed solution is an implementation of Tito Orlandi's concept of the 'table of signs'²¹. Namely, I am pairing the encoding of each witness with a specific table of graphemes and a specific table of alphabemes. This will be discussed in paragraph 9.4.1. *A Saussurean solution: the tables of signs*.

10.3. The 'musical score' model

10.3.1. The abstract data model

As mentioned above, my edition model resembles a musical score, with three parallel transcriptions of the text (graphic, alphabetic and linguistic) mapped to one another at the granularity level of single graphemes.

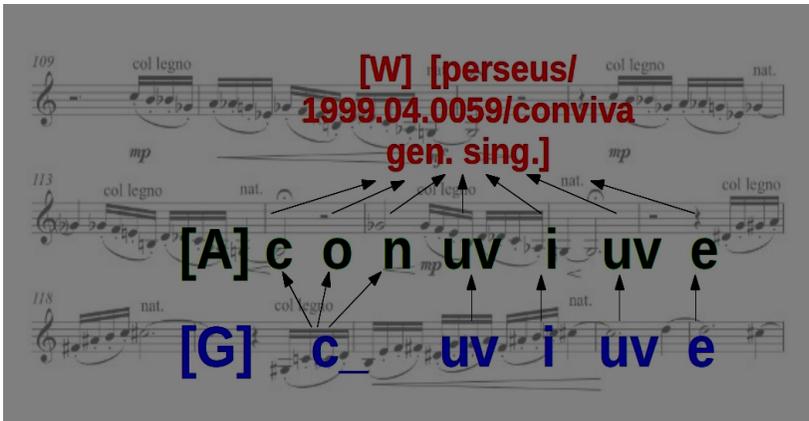


Fig. 10.2. The 'musical score' model. [G] stands for graphical layer, [A] for alphabetic layer and [W] for linguistic layer (a sequence of inflected words).

²¹ See Orlandi 2010, 38-43.

This edition model is approximated in Fig. 10.2²². In this example, grapheme 'c_' at the graphic layer (identifying a systematic abbreviation for 'con') is mapped to three alphabemes ('c', 'o' and, 'n') at the alphabetic layer. The manuscript hypothetically encoded here does not distinguish between the 'u' and 'v' alphabemes: it has a unique alphabeme with digital ID 'uv' (and a unique corresponding grapheme, whose ID is also 'uv').

The word is the genitive singular of 'conviva' ("table companion"). Note that today this inflected word would be spelled 'convivae'. Its alphabetic transcription in Fig. 10.2 does not constitute a 'regularised spelling'. It is simply a sequence of alphabemes 'as they are' in the manuscript²³, reflecting the medieval spelling 'convivae'.

10.3.2. The linguistic layer: inflected words as atoms

The encoding units of the linguistic layer ([W] in Fig. 10.2) are inflected words. In Orlandi's view, each inflected word should not be represented as a sequence of letters, a simplistic solution that poses the issues of homographs and variant spellings, but by means of a unique digital identifier. This corresponds conceptually to a 'table of signs' for the linguistic layer, but clearly the attribution of an identifier to many thousands of words poses two kinds of issues. These are the conceptual and the practical:

Conceptual issues. If we do not identify inflected words by means of a sequence of digital characters (i.e. on the basis of alphabetic writing), how else can we identify them? A word might be identified by a combination of a lemma (with a unique ID retrieved from a standard digital vocabulary) and standardised morphological information. Therefore, Latin 'deum' will no longer be identified simply by a sequence of four Unicode code points, which would not tell us whether it is the accusative singular of 'deus' ('god') or the archaic form of its genitive plural (classical Latin 'deorum'). It would instead be encoded essentially as a combination of the following two pieces of information:

²² The approximation lies in the fact that the elements of the graphic layer [G] should also be mapped to the linguistic layer [W].

²³ i.e. as the philologist interpretively extracts them from the graphic encoding of the manuscript.

- The identifier of lemma 'deus' ('god') in Perseus' Lewis-Short dictionary: perseus/1999.04.0059/deus;
- A standardised encoding of: 'genitive, plural'.

It might be necessary, however, to add a third piece of information: the spelling of the word, i.e. the sequence of the four Unicode code points for 'deum'. This would allow us to distinguish between 'deum' and 'deorum', which otherwise would erroneously be encoded in the same way (lemma: 'deus'; morphology: genitive plural)²⁴.

Practical issues. It is obviously easier to identify inflected words by a sequence of 'characters', either keyed in by hand or obtained by OCR. Encoding every single word of every manuscript as a combination of a lemma and some morphological information implies that the philologist has to perform lemmatisation and morphological parsing on every word. The only solution is to make both operations semi-automatic with the help of dedicated software that makes the work-flow sustainable, though undoubtedly more burdensome for the encoder. This will be discussed in paragraph 9.5.1. *Current work-flow sotto.*

10.3.3. Two possible XML linearisations of the 'musical score' data model

Let us go back to the 'musical score' data model and its possible linearisation. XML appears to be better suited to represent trees than three sequences of tokens flowing in parallel and aligned with each other at such a level of granularity. Therefore, I am still open to the option of turning to other text-encoding data models, including the range-based model elaborated by Gregor Middel and others within the Faust Project²⁵, Desmond Schmidt's multi-version

²⁴ This approach may look like a simple addition of lemmatisation and morphological parsing to a regular Unicode string ('deus'), and therefore an unnecessary complication. A key question in this respect is whether there exists such a thing as a homographic variant, i.e. whether two witnesses can bear two sequences of signs which resolve to the same alphabetic sequence but can be reasonably interpreted as different readings, due to some contextual information. There is no room for such a discussion here: see my talk Monella 2012, slides 42-48.

²⁵ The Faust Project is also elaborating a data model in which documents are encoded at more than one level, namely two: "dokumentarische" and "textuelle Transkript" (the latter is similar to my linguistic layer). See Bohnenkamp et al. 2011, especially section III. *Umsetzung* (pp. 38-65) and Brüning, Henzel, Pravida 2012.

documents²⁶ and Manfred's Thaller's extended strings²⁷.

At the moment, however, I am still experimenting on TEI/XML to see whether I can adapt it to serve the purposes of my edition model. My work on this is still in progress, so what follows is a mere working hypothesis. I shall now describe the two XML linearisation models I am currently testing:

1. Separate files linearisation model
2. Menota linearisation model

10.3.4. Separate files linearisation model

This model requires that, for every witness, three different TEI P5 'XML transcription files' are created²⁸:

1. `salm_graphic.xml`
2. `salm_alphabetic.xml`
3. `salm_linguistic.xml`

`salm_graphic.xml`. This is the transcription of the text of the *Codex Salmasianus* at the graphic layer. It includes a sequence of TEI P5 `<g>` elements representing graphemes, paragraphematic signs (like punctuation) and other graphic signs (like spaces between words).

The following code is taken from the current graphic XML transcription file of the *Codex Salmasianus* and represents the first two words of the first line of the poem ('Ter ternae'):

```
<g id="10.1" ref="#T" />
<g id="10.2" ref="#e" />
<g id="10.3" ref="#r" />
<g id="11.1" ref="#t" />
<g id="11.2" ref="#e" />
<g id="11.3" ref="#r" />
<g id="11.4" ref="#n" />
<g id="11.5" ref="#ae" />
```

²⁶ Schmidt & Colomb 2009; Schmidt & Fiormonte 2010.

²⁷ Thaller 2006.

²⁸ The 'salm_' prefix of filenames denotes that they refer to the encoding of *Codex Salmasianus*, the first witness I am encoding.

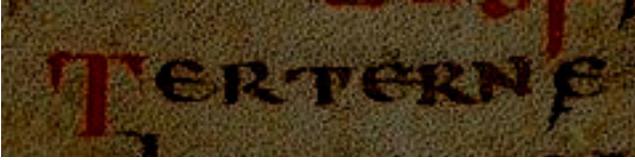


Fig. 10.3. The first two words of the first line of the *Iudicium coci et pistoris* in the *Codex Salmasianus*: 'Ter ternae'.

The @id attributes have the usual function of marking each grapheme unambiguously and sequentially. For readability's sake, the value of @id is the number of the word ('Ter' is the 10th word of the transcription, as it is preceded by the words of the title line), followed by a dot and the number of the grapheme within the word. The manuscript has no spaces between words (*scriptio continua*), so the delimitation of a sequence of graphemes as a 'word' is an interpretive act of the philologist. Also note that, as there are no graphic signs (spaces) to mark the distinction between words, in the graphic transcription above there is no element marking the distinction between the two words. Line breaks between verses, instead, are reported in this file because although they are not ink marks, they are still graphic signs found in the document.

Grapheme 10.1 above is an uppercase Latin 'T': as it is a grapheme (not an alphabeme), it is distinct from a lowercase Latin 't'. Grapheme 11.5 (the last sign on the right in Fig. 10.3) has the two alphabemes 'a' and 'e' as alphabetic content, and 'ae' as digital identifier. I shall turn back to the semantics of the @ref attributes and on the use of <g> elements in paragraph 9.4.2. *The table of graphemes* below.

salm_alphabetic.xml. This is the transcription of the same witness at the alphabetic layer. It includes a sequence of TEI P5 <g> elements which, in this file, represent alphabemes. What follows is a sampling of the current salm_alphabetic.xml file:

```
<g id="10.1.1" ref="#t" />
<g id="10.2.1" ref="#e" />
<g id="10.3.1" ref="#r" />
<g id="11.1.1" ref="#t" />
<g id="11.2.1" ref="#e" />
<g id="11.3.1" ref="#r" />
<g id="11.4.1" ref="#n" />
<g id="11.5.1" ref="#a" />
<g id="11.5.2" ref="#e" />
```

In fact, the introduction of a new `<alphabeme>` element would theoretically be a much better solution. This is still an option on the table. However, for the time being I am using the existing `<g>` element while giving it different semantics than in file `salm_graphic.xml`. Also the value of `@ref` attributes in the code above has different semantics, as will be discussed in paragraph 9.4.3. *The table of alphabemes* below.

In the value of the `@id` element of `alphabeme 11.5.1` above, the three numbers mean that this is the 1st *alphabeme* ('a') represented by the 5th *grapheme* (the abbreviation for diphthong 'ae') of the 11th *word* ('ternae').

`salm_linguistic.xml`. This is the transcription of the same witness at the linguistic layer. It includes a sequence of TEI P5 `<w>` ('word') empty elements. They are empty because words, in this model, are not represented by a sequence of alphabemes. This is the part of the model that still requires most work. The following code simply aims to give an idea of the general concept:

```
<w id="10" ana="adv, [ter], ter" />
<w id="11" ana="adj, [ternus], n, p, f, ternaes" />
```

At the current stage of development of the model, a generic `@ana` attribute includes the digital identifier of the inflected word: in the example above, word 11 is the *nominative plural feminine* of lemma 'ternus' (which is an *adjective*), whose contemporary standard spelling is 'ternae'. At this stage, this notation is nothing more than a placeholder. At a later stage in the development of the prototype, this information will be distributed into the relevant analytical attributes, like `@type`, `@lemma`, `@lemmaRef` etc.²⁹

The elements of the three files above (graphemes, alphabemes and inflected words) must be aligned with each other. This task is performed by three more 'alignment files':

1. `salm_align_alph_graph.xml`
2. `salm_align_alph_ling.xml`
3. `salm_align_graph_ling.xml`

These files include `<link>` and `<ptr>` elements only. Their function

²⁹ Section 17 *Simple Analytic Mechanisms* of the TEI P5 Guidelines (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html>, last retrieved 18/03/2013).

is to perform the mapping described in Fig. 10.4. I shall now briefly describe how the alignment information is encoded in each file.

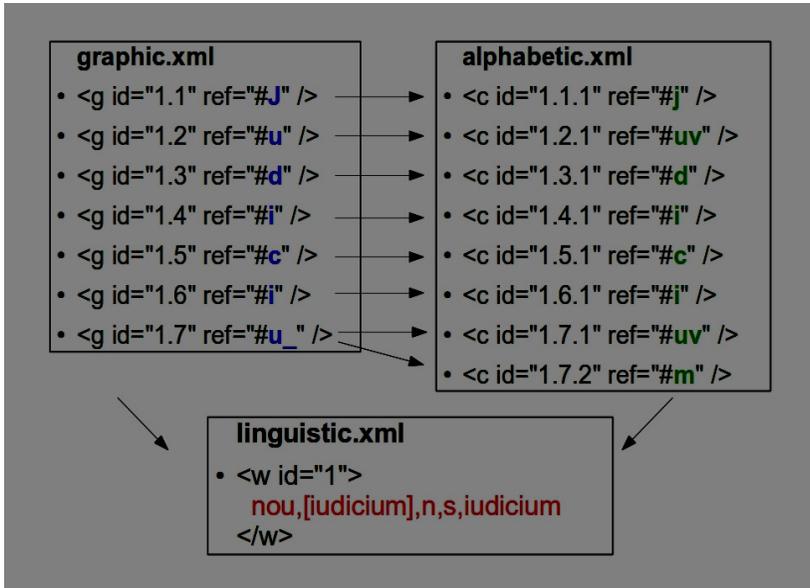


Fig. 10.4. Alignment between the three separate XML transcription files

salm_align_alph_graph.xml. There must be a formal way to encode the information so that, for example, grapheme 11.5 in salm_graphic.xml (abbreviation 'ae') corresponds to alphabemes 11.5.1 ('a') and 11.5.2 ('e') in salm_alphabetic.xml. This function is performed by a number of TEI P5 `<link>` elements stored in a separate file. The XML file aligning salm_alphabetic.xml and salm_graphic.xml is currently named salm_align_alph_graph.xml. This is a portion of its content:

```
<link targets="salm_graphic.xml#11.5 #11.5" />
<ptr id="11.5" targets=
    "salm_alphabetic.xml#11.5.1
    salm_alphabetic.xml#11.5.2" />
```

The `<ptr>` element is needed to encode a one-to-many link in TEI P5 (one grapheme must here point to two alphabemes)³⁰.

salm_align_alph_ling.xml. This file aligns the alphabetic tran-

³⁰ Section 16.1.4 *Intermediate Pointers* of the TEI P5 Guidelines (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SAPTIP>, last retrieved 18/03/2013).

scription with the linguistic one, i.e. an inflected word with a sequence of alphabemes that represent it. This is the snippet of code that aligns word 11 ('ternae') with the corresponding alphabemes:

```
<link targets="salm_linguistic.xml#11 #11" />
<ptr id="11" targets=
    "salm_alphabetic.xml#11.1.1
    salm_alphabetic.xml#11.2.1
    salm_alphabetic.xml#11.3.1
    salm_alphabetic.xml#11.4.1
    salm_alphabetic.xml#11.5.1
    salm_alphabetic.xml#11.5.2" />
```

salm_align_graph_ling.xml. This file aligns the graphic transcription with the linguistic one. The following code aligns the same word ('ternae') with the *graphemes* that encode it. Note that, while the word has 6 alphabemes (see code above), it is here linked to *only 5* graphemes, as the final diphthong 'ae' is one grapheme in the manuscript:

```
<link targets="salm_linguistic.xml#11 #11" />
<ptr id="11" targets=
    "salm_graphic.xml#11.1
    salm_graphic.xml#11.2
    salm_graphic.xml#11.3
    salm_graphic.xml#11.4
    salm_graphic.xml#11.5" />
```

The source code above does look complex at first sight, but of course none of it is written directly by the philologist. The writing of the actual XML files, the attribution of @id numbers and the simultaneous creation of the relevant links between graphemes, alphabemes and inflected words are all tasks performed by a small piece of software (currently a Python 3.3 script, `input.py`). What the philologist actually writes is a simple 'input transcription file' in CSV format (`salmasianus.csv`, for Codex Salmasianus). The script inputs this file and outputs the three XML transcription files (`salm_graphic.xml`, `salm_alphabetic.xml`, `salm_linguistic.xml`) and the three 'alignment files' (which pair each transcription with the other two). This will be discussed in detail in paragraph 9.5.1. *Current work-flow* below.

All XML, CSV and Python files described in this article are openly available in the GitHub repository <https://github.com/paolomonella>

[/vespa.git](http://vespa.git).

10.3.5. Menota linearisation model

While I was in search of an already existing customisation of TEI that could encode the text at different parallel layers, my attention was drawn by Roberto Rosselli Del Turco to The Medieval Nordic Text Archive (Menota)³¹. The project's customisation of TEI provides a mechanism to encode a text at three layers, named "facsimile", "diplomatic" and "normalised" (which, respectively, correspond roughly to my graphical, alphabetic and linguistic layers). To encode the text in this way, the Menota Project added three elements to the TEI/XML schema, namely `<me:fac>`, `<me:dipl>` and `<me:norm>`³². The resulting code looks like this:

```
<w>
  <choice>

  <me:fac>&drot;<am>&osup;</am>ttin<am>&bar;</am></m
e:fac>
    <me:dipl>d<ex>ro</ex>ttin<ex>n</ex></me:dipl>
    <me:norm>Dróttinn</me:norm>
  </choice>
</w>
```

The Menota encoding practice for primary sources is certainly a good basis to experiment upon, but it differs from my envisioned model in four ways:

1. *Granularity*. The finest granularity of the current Menota markup is at word level, while I need alignment at grapheme-level granularity.
2. *Grapheme/alphabeme distinction*. All three Menota transcription layers share the same set of 'characters', while my graphic and alphabetic transcriptions are based on different sets of elements (graphemes and alphabemes respectively).
3. *Unicode-compliance principle*. Menota relies on Unicode for the definition of each sign encoded, while I want to have a formal and explicit definition of each element (grapheme or alphabeme) used

³¹ Home page: http://www.menota.org/EN_forside.xhtml (last retrieved 18.03.2013). See Haugen 2004.

³² Paragraph 3.2 Levels of text representation in The Menota Handbook v 2.0: http://www.menota.org/HB2_ch3.xml#d481e406 (last retrieved 18.03.2013).

in the transcription (see paragraph 9.2.2. *A Saussurean issue sopra*).

4. *Linguistic transcription*. My linguistic transcription layer should not encode inflected words as a sequence of letters, as Menota does, but by means of unique digital identifiers, as described in paragraph 9.3.2. *The linguistic layer: inflected words as atoms sopra*.

Still, the Menota customisation of TEI P5 has the clear advantage of allowing the coexistence of three different layers of transcription in the same XML file. The following code illustrates a hypothetical combination of the Menota 'three-layers' innovation and my own proposed encoding of graphemes, alphabemes and inflected words:

```
<w id="11">
  <me:facs>
    <g id="11.1" ref="#grapheme_t" />
    <g id="11.2" ref="#grapheme_e" />
    <g id="11.3" ref="#grapheme_r" />
    <g id="11.4" ref="#grapheme_n" />
    <g id="11.5" ref="#grapheme_ae" />
  </me:facs>
  <me:dipl>
    <g id="11.1.1" ref="#alphabeme_t" />
    <g id="11.2.1" ref="#alphabeme_e" />
    <g id="11.3.1" ref="#alphabeme_r" />
    <g id="11.4.1" ref="#alphabeme_n" />
    <g id="11.5.1" ref="#alphabeme_a" />
    <g id="11.5.2" ref="#alphabeme_e" />
  </me:dipl>
  <me:norm>
    adj, [ternus], n, p, f, ternae
  </me:norm>
  <link targets="#11.1 #ptr11.1" />
    <ptr id="ptr11.1" targets=
      "#11.1.1" />
  <link targets="#11.2 #ptr11.2" />
    <ptr id="ptr11.2" targets=
      "#11.2.1" />
  <link targets="#11.3 #ptr11.3" />
    <ptr id="ptr11.3" targets=
      "#11.3.1" />
  <link targets="#11.4 #ptr11.4" />
```

```

    <ptr id="ptr11.4" targets=
      "#11.4.1" />
  <link targets="#11.5 #ptr11.5" />
  <ptr id="ptr11.5" targets=
    "#11.5.1
    #11.5.2" />
</w>

```

Of course, this depends on the possibility of allowing, under the Menota scheme, a hierarchy like `<w>` / `<choice>` / `<me:fac>` (or `<me:dipl>`, or `<me:norm>`) / `<g>`³³.

As word/graphemes and word/alphabemes alignment is granted by the inclusion of Menota `<me:fac>` and `<me:dipl>` elements in the same `<w>` element, `<link>` elements are only required for graphemes/alphabemes alignment. `<link>` elements are now included in the same `salm_menota.xml` file.

In this hypothesis, issues 1 and 2 above (granularity and grapheme/alphabeme distinction) are overcome by the pervasive use of `<g>` elements, aligned through `<link>` elements. Issue 4 above (linguistic transcription) may be overcome by the introduction of non-alphabetic representation of inflected words in the `<me:norm>` element. My proposed strategy for the solution of issue 3 above (Unicode-compliance principle) is centred on the use of the `<g>` element and can be applied either to the 'Separate files linearisation model' or to the 'Menota linearisation model' equally. This strategy will be described in detail in the following paragraphs.

³³ Some of the proposed modifications to the Menota encoding practice might require a further customisation of the Menota scheme.

10.4. Tables of signs

10.4.1. A Saussurean solution: the tables of signs

To solve the 'Unicode-compliance' issue discussed in paragraph. 9.2.2. *A Saussurean issue* sopra, I propose a TEI/XML implementation of Tito Orlandi's idea of "tabella dei segni" (henceforth 'table of signs')³⁴.

In Orlandi's view, when a philologist digitally encodes the 'glyphs' of a manuscript or any other set of textual phenomena in a document, he or she must create an 'ideal' table (that can, however, become an actual table as part of an edition's documentation), the 'left column' of which consists of the digital signs that they use to represent the textual phenomena, while the 'right column' "includes the list of single phenomena of the object of the encoding which one wants to encode". Orlandi's formulation is conceptually wide, and intentionally does not specify the nature of the textual phenomena of the 'right column'³⁵.

In this respect, I am proposing two innovations:

1. I am creating two different tables of signs: the table of graphemes and the table of alphabemes;
2. I am looking for a suitable format to include the two tables in my digital edition as a formal, computable and stable component of the XML source code, possibly in the `<charDecl>` section of the TEI header.

10.4.2. The table of graphemes

In the current stage of development of my edition, this is a simple CSV file (`salm_table_graphemes.csv`) with three columns, which my Python script (`input.py`) transforms into TEI/XML code and writes in the TEI header of the `salm_graphic.xml` and `salm_menota.xml` transcription files. Tab. 1 below shows several rows from my current table of graphemes for the *Codex Salmasianus*:

³⁴ Orlandi 2010, 38-43.

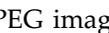
³⁵ The quotation is from Orlandi 2010, 38 (the translation is mine). For his edition of the *Zibaldone Laurenziano* (see footnote 3 above), Raul Mordenti is creating a table of signs describing the set of glyphs identified in the manuscript ('right column', textual phenomena), mapped to a set of digital signs, i.e. ASCII characters or XML entities in the form `&abc;` ('left column'). This is an actual table meant to be published with the documentation of the digital edition (Mordenti 2012).

Digital ID	Content of the grapheme (=alphabeme[s])	Expression of the grapheme (=glyph)	Visualisation
	uv	u/v Latin minuscule uncial (u-shaped, not v-shaped)	u
z	z	Latin minuscule uncial z	z
ae	a,e	Latin minuscule uncial e with a tail bottom left, 	ε

Tab. 10.1. A portion of the table of graphemes for the *Codex Salmasianus*.

Again, note that this file and the others mentioned in this article (and available in <https://github.com/paolomonella/vespa.git>) are still at a prototype stage and are merely reported to give a practical idea of the path that my experimentation is currently following.

The 'Digital ID' column includes strings of Unicode characters, but they might just as well be ASCII characters or numerals. They unambiguously identify every sign (in this case, every grapheme) and correspond to Orlandi's 'left column'³⁶.

For the other two columns ('Grapheme: content' and 'Grapheme: expression'), a premise must be made: all digital elements upon which our digital textual encoding is based (Unicode characters, XML entities, <g> elements etc.) are signs. As such, they have an expression (e.g. a Unicode code point) and content. A key concept of my edition is that the *content* of such digital signs is *another sign* (e.g. a grapheme), which, in its turn, is constituted and defined by another expression/content pair. For example, in the third row of the table above, the digital sign identified by the two Unicode characters 'ae' has these two digital characters as its expression, and the corresponding grapheme as its content. In its turn, this grapheme consists of the pairing of its expression (the 'e with a tail' glyph described in the third column and visible in the JPEG image ) with its content (the two alphabemes 'a' and 'e', listed in the second column). Thus Orlandi's 'left column', describing the textual phenomenon, is here represented by *two* columns (the second and third in table 1 above).

The fourth column ('Visualisation') has an eminently practical function: it instructs visualisation software (XSLT or other technology) on what Unicode character(s) it should use to display that grapheme,

³⁶ Orlandi 2010, 38.

if it is required to give a screen or print visualisation of the graphic layer of the text.

10.4.3. The table of alphabemes

This too is currently a simple CSV table (`salm_table_alphabemes.csv`), which the `input.py` Python script transforms into XML code. The table for the *Codex Salmasianus* has 20 rows, as this is the number of alphabemes that I identified in the codex³⁷. This is a part of the table:

Digital ID	Description of the Alphabeme	Visualisation
t	Latin t	t
uv	Latin u/v	u
z	Latin z	z

Tab. 10.2. A portion of the table of alphabemes for the *Codex Salmasianus*.

There is a number of issues connected with this table which are still open. Alphabet and alphabemes are conceptual objects much more complex than they appear at first sight³⁸.

In the tentative Tab. 2 above, I did not describe alphabemes by means of an expression/content pair, simply because I am not sure of what that would be. One might think that the expression of an alphabeme is its corresponding grapheme, and its content is a phoneme. The very invention of the alphabet had the goal of making things just this simple. However, things are not so simple, because alphabets are cultural objects in themselves³⁹.

1. *Is a phoneme the content of an alphabeme?* What is the phonetic content of alphabeme 'Latin letter c' in *Codex Salmasianus*, a Latin document written in the VII or VIII Century CE in Vandalic Africa, where 'letter c' was probably pronounced in different ways (not fully known to us) according to the vowel that followed, and to the social and cultural status, geographic origin and ethnicity of the reader? Not to mention that the same 'Latin letter c' in the same text has

³⁷ The second of the three rows in Tab. 2 shows that the *Codex Salmasianus* does not have two different alphabemes 'u' and 'v', but one only: this is why there are 20 rows in the table instead of 21, a more familiar number for the Latin alphabet.

³⁸ Mordenti 2011, 640-648 has an interesting reflection on such complexity.

³⁹ Suffice it to mention Sampson 1990.

been read in many different ways throughout the centuries by the many readers that happened to have that codex in their hands. In other words, 'dead languages' have the advantage of keeping alive in us a critical sense of the correspondence between alphabemes and phonemes. But even in contemporary languages, the same English 'letter' may be connected to slightly different phonemes in different national, regional, dialectal or socially determined pronunciations. This should already suffice to show that an alphabeme is defined *per se* as a cultural object, not as a content/expression (phoneme/grapheme) pair.

2. *Is a grapheme the expression of an alphabeme?* This hypothesis is conceptually less problematic, but is still rather simplistic. For instance, grapheme 'b' may represent alphabeme 'b', but grapheme 'b_' in Codex Salmasianus (a 'b' with a macron top right, serving as abbreviation for 'bis') also represents alphabeme 'b', as well as alphabemes 'i' and 's'.

I would still be ready to create a column in the table of alphabemes where grapheme 'b' is designated as the *main* expression (though this is a rather loose concept) of alphabeme 'b', but I would hesitate to create a column including the contents of alphabemes, as I currently would not know how to populate it.

The last column on the right in the table above, again, instructs the software on what Unicode character may be best used in visualising of the alphabetic layer of the text.

Lastly, it is important to note that the second column of the table of *graphemes* (Tab. 1) points to the IDs (first column) of the table of *alphabemes* (Tab. 2).

10.4.4. Inclusion of the tables of signs in the TEI header

As anticipated, it is my intention to make the two tables of signs an integral part of the source code in the XML transcription files. This is, however, very problematic and requires further work. Again, I shall here describe the hypotheses that I am currently working on.

If we adopt the 'separate files linearisation model' described in paragraph 9.3.4. *sopra*, in the *Codex Salmasianus* the table of graphemes should be integrated in file `salm_graphic.xml` and the table of alphabemes in file `salm_alphabetic.xml`. The most obvious place for these tables is the `<charDecl>` element in the TEI header.

TEI P5 features an interesting innovation: the philologist can define 'non-standard characters' in the `<charDecl>` element of the header, and then encode instances of them in the body by means of `<g>` elements. This appears to be a more sophisticated mechanism for formally defining a grapheme than XML entities, so I decided to adopt it in my edition.

A major problem, however, is that to my knowledge, there is no way in TEI to formally define simple Unicode characters.

In other words, if I encode something like `ui<g ref="ae" />` (for Latin 'viae' written with a final 'æ' grapheme), I can formally define the last grapheme (the 'non-standard' 'æ') in `<charDecl>`, but I have no way to define the first grapheme (the 'standard' 'u', for which I am using Unicode code point U+0075) anywhere. This is because the TEI Unicode-compliance principle described in 9.2.2. *A Saussurean issue* sopra assumes that a 'u' is a 'u', and requires no further definition than the *absolute* one given by the Unicode standard. However, as my 'Saussurean' argument in paragraph 9.2.2. sopra and the very definition of grapheme in the *Glossary of Unicode Terms* postulate, a grapheme must be always defined in a *relative* way, "in the context of a particular writing system"⁴⁰. In the context of a manuscript featuring a 'u'/'v' distinction, a 'u' is *not the same thing* as a 'u' in the context of a manuscript with no such distinction. This implies that it must always be possible – in fact, that it should be required – to give a description of *all* graphemes identified in a manuscript, both 'standard' and 'non-standard'. In the example above, not only the final `<g ref="ae" />`, but also the initial Unicode character U+0075 (for 'u') should be matched by a (brief and formal) description in `<charDecl>`.

Of course, all that is being said about the need to define all graphemes also applies to alphabemes, allographs and the other kinds of signs that the philologist decides to encode.

At the moment, as I said, it is technically not possible to formally define Unicode characters, and I am afraid that such a change would require not only much work on the TEI schema, but also a quite radical modification in the approach towards 'characters' in the Guidelines, in which

⁴⁰ The quotation here comes from the definition of "grapheme" in the *Glossary of Unicode Terms* (see footnote 18 above). The need for the 'local' definition of signs is postulated by Orlandi 2010, 38-43 (but see also pages 9 and 48).

1. it should be recommended to define *all* graphemes (and other signs), not ;
2. the very conceptual distinction between 'standard' and 'non-standard' characters should be eliminated.

As this seems to be more work than I am currently willing to undertake, for the time being I am confining myself to using only `<g>` elements (not Unicode characters) in the body of the XML transcription files, as I know that I can define them in the `<charDecl>`. I hope that a solution for this issue can be found in the future, allowing the philologist to simply encode `ui<g ref="ae" />` rather than

```
<g ref="uv" />
<g ref="i" />
<g ref="ae" />
```

Needless to say, it would be utterly infeasible to key the 'all `<g>`' source code above directly. This is another aspect for which the `input.py` Python script is of use. What I am currently keying is something like this: `ui(ae)`. The script transforms this into the sequence of three `<g>` elements above. This will be further discussed in paragraph 9.5.1. *Current work-flow* sotto.

The following code is taken from file `salm_graphic.xml`. It exemplifies the initial description of graphemes in `<charDecl>` and their subsequent use in `<body>`:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <!-- ... -->
<charDecl>
  <!-- ... -->
  <char xml:id= "t">
    <charProp>
      <localName>Expression</localName>
      <value>Latin minuscule uncial t</value>
    </charProp>
    <charProp>
      <localName>Content</localName>
      <value>t</value>
    </charProp>
```

```

        <charProp>
            <localName>Visualisation</localName>
            <value>t</value>
        </charProp>
    </char>
    <!-- ... -->
</charDecl>
</teiHeader>
<text>
<body>
    <!-- ... -->
    <g id="24.1" ref="#e" />
    <g id="24.2" ref="#t" />
</text>
</body>
</TEI>

```

Likewise, alphabemes are described in the `<charDecl>` section of file `salm_alphabetic.xml`:

```

<char xml:id= "uv">
    <charName>Latin u/v</charName>
    <desc>Latin alphabeme 'u', with no 'u'/'v'
opposition</desc>
    <charProp>
        <localName>Visualisation</localName>
        <value>u</value>
    </charProp>
</char>

```

Note that the less structured description method for alphabemes (`<desc>`) mirrors the simpler current structure of the table of alphabemes (see Tab. 2 above).

If we adopt the 'Menota linearisation model' described in paragraph 9.3.5., in addition to the issues already discussed, a further complication arises: both tables (graphemes and alphabemes) should be included in the same `salm_menota.xml` file. Unfortunately, the TEI header is not designed to accommodate two different `<charDecl>` elements marked as belonging to two different textual layers. This too would require an *ad hoc* customisation of TEI.

10.4.5. Collation

As anticipated at the beginning of this article, my edition is still at an early stage, namely that of encoding witnesses, a stage which runs parallel to the elaboration of experimental encoding standards. The collation phase is yet to come, but I can here mention the two principles that will lead it.

1. *Scribes and scholars*. I mean to encode the text of modern and contemporary editions, and also scholarly editions, alongside that of ancient medieval manuscripts, thus making no artificial distinction between 'scribes' and 'scholars'. For the same principle, I shall produce 'my' text, by comparing the extant variants and my own *iudicium*, but this text will then be presented to the reader at the same level as any other text from ancient or modern witnesses⁴¹.
2. *Collation performed layer by layer*. The *linguistic layer* of one witness will be collated with the same layer of others. This should be somewhat easier, as all witnesses' texts will be encoded at this layer by means of references to the same dictionary (for lemmas) and to the same formalised grammar (for morphology)⁴².

Collating the *alphabetic layer* of a manuscript with the alphabetic layer of a modern print edition, however, will be more complicated, as two witnesses may well not share exactly the same alphabet (e.g. for the 'u' / 'v' distinction issue). Still, this problem can be solved with a formal table of alphabemes for each manuscript describing each reference alphabet. This might allow us to create, by means of RDF/OWL or other standards, a net of correspondences: for example, the 'uv' alphabeme in MS B might be connected to the 'u' and the 'v' alphabemes in MS A. This should suffice to instruct sophisticated

⁴¹ I shall mention only some of the main reading that influenced me in this respect: Pasquali 1971; Reynolds & Wilson 1991 (whose title is "Scribes and scholars"); Cozzo 2006; Benozzo 2011; Fiormonte 2012b.

⁴² The most obvious candidate for a reference dictionary is the online version of the *Latin Dictionary* by Charlton T. Lewis and Charles Short in the *Perseus Digital Library* (<http://www.perseus.tufts.edu/hopper>, last retrieved 19.03.2013). A theoretical issue regarding lemmata and morphology is that both the vocabulary and the grammar of Latin evolved over the centuries, and one could argue that Virgil's text in a medieval manuscript should be mapped against the vocabulary and grammar of the manuscript's time – therefore not those of 'classical' Latin. This objection, however, can be partially rejected on the basis that the scribes/philologists normally meant to preserve the text in a presumed 'original' linguistic form.

collation software to compare two witnesses at this layer.

The very same issue will affect collation at the *graphic layer*, but on a much larger scale: just think of the enormous variance in the nature and use of punctuation in manuscripts. I am considering the option of skipping collation at this layer altogether – or alternatively adopting the same approach that I outlined for the alphabetic layer.

10.5. Work-flow

10.5.1. Current work-flow

I shall lastly describe the current work-flow of my edition. As said before, all working files described here are available in the the GitHub repository <https://github.com/paolomonella/vespa.git>, while further documentation and discussion on the open issues will be published in <http://www.unipa.it/paolo.monella/lincei/edition.html> as the work proceeds.

I am currently encoding the oldest manuscripts. What I actually edit 'by hand' is a CSV file for each manuscript. This is a snippet of the content of file `salmasianus.csv`, the current CSV transcription file for the *Codex Salmasianus* ('\$' is the CSV delimiter character):

Ter	\$	adv, [ter], ter
tern(ae)	\$	adj, [ternus], n, p, f, ternae
uarias	\$	adj, [varius], ac, p, f, varias

The first column includes the graphic transcription. This is what I actually key:

- If the grapheme's ID is composed of one Unicode character (as for uppercase 'T' in 'Ter', above), I simply key that character inline.
- If the grapheme's ID is composed of two or more characters (as for the 'ae' grapheme in 'ternæ'), I still key it inline, but wrapped in parentheses, so the script `input.py` knows how to process it.

Though the *Codex Salmasianus* has no regular graphic word distinction (no spaces between words), I am interpretively distinguishing words: in each row, the left cell has a sequence of graphemes mapped to an inflected word in the right cell. For the time being, this is a sufficiently practical and efficient way to encode the graphemes/word mapping with a simple text or spreadsheet editor.

The script `input.py` does the following (I am using here, as elsewhere in this article, the files relative to *Codex Salmasianus* as an example):

1. *Importing files.* It imports the tables of graphemes and alphabemes (`salm_table_graphemes.csv` and `salm_table_alphabemes.csv`) and the CSV transcription file (`salmasianus.csv`) of the witness.
2. *Tables of signs.* It converts the two tables of signs into a sequence of `<char>` elements as described above and writes them in the `<charDecl>` section of the `<teiHeader>` of files `salm_graphic.xml`, `salm_alphabetic.xml` and `salm_menota.xml`.
3. *Graphic transcription.* For each row of the transcription CSV file (i.e. for each word), it processes the 'left column' strings (i.e. the graphic transcription) and writes the 'all `<g>`' XML code described above to files `salm_graphic.xml` and `salm_menota.xml`. In the latter file, the script inserts this sequence of `<g>` elements in a `<me:fac>` element.
4. *Alphabetic transcription.* Each 'left column' string is a sequence of grapheme IDs. The script matches these IDs in `table_graphemes.csv` with the ID(s) of the corresponding alphabeme(s). For instance, grapheme 'ae' in `table_graphemes.csv` corresponds to the sequence of alphabemes 'a', 'e'. The script writes the relative `<g>` elements (representing alphabemes) to files `salm_alphabetic.xml` and `salm_menota.xml`. In the latter file, the script inserts this sequence of `<g>` elements in a `<me:dipl>` element.
5. *Linguistic transcription.* The script processes the 'right column' (linguistic transcription) of the CSV transcription file and writes its content to `salm_linguistic.xml` (with a `<w>` element), and to `salm_menota.xml` (in a `<me:norm>` element).

As far as the graphic and alphabetic encoding is concerned, this work-flow is very efficient. Apart from the tables of signs, what I actually key in is the graphic transcription alone, in a sort of 'dialect format' for internal use: e.g. `tern(ae)` for 'ternæ'. The alphabetic transcription is generated automatically.

10.5.2. Prospective work-flow

At the present stage, the 'missing link' in this work-flow is the linguistic layer. In fact, I currently populate the 'right column' of the CSV transcription file by hand. I do so simply because I do not want to leave those cells blank. As anticipated, however, I plan to develop

an efficient input mechanism for the linguistic transcription (possibly by means of a web interface and JavaScript) that will run as follows:

1. The philologist keys in the graphic transcription in a dynamic HTML page in the current 'dialect format', e.g. `tern(ae)`.
2. JavaScript generates the alphabetic transcription layer semi-automatically (the philologist approves it or rejects it).
3. A 'normalised' string resulting from the processing of the alphabetic transcription⁴³ is sent to a web service (possibly *Perseus*⁴⁴) that parses and lemmatises it, thus returning the lemmatic and morphological information needed for the linguistic layer encoding. This process will also be semi-automatic, as in most cases *Perseus* will return a number of possible lemmas or morphological analyses, between which the philologist will be asked to choose. The whole process should be performed by JavaScript and other HTML5 technologies.

Undoubtedly, this envisioned work-flow will require more work on the side of the philologist than a simple 'one-layer' transcription, but I think that it may be considered a sustainable work-flow when compared to the complexity of the edition model that it would produce.

10.6. Conclusion

I believe that a sophisticated, multi-layered model requiring a complete description of encoded signs is needed to meet the challenge of creating digital scholarly editions which are based on the collation of textual witnesses written with different graphic and alphabetic systems.

I hope that I shall soon be able to develop a complete and sustainable work-flow that will allow me to implement such a model in a working prototype, and to submit the results of this experiment to the attention of the Digital Humanities community.

⁴³ This is one of the phases in which the values of the 'visualisation' field in the table of alphabemes come of use.

⁴⁴ See footnote 41.

10.7. Bibliography

- BARONI A. (2009). *La grafematica: teorie, problemi e applicazioni*, Master's thesis, Università di Padova. URL= http://unipd.academia.edu/AntonioBaroni/Papers/455456/La_grafematica_teorie_problemi_e_applicazioni [last retrived 10.03.2013].
- BARRY B. (1987). *The Iudicium coci et pistoris of Vespa*, in *Filologia e forme letterarie. Studi offerti a Francesco della Corte*, vol. 4, Quattro Venti, pp. 135-149.
- BAUMGARTNER A. (1981). *Untersuchungen zur Anthologie des Codex Salmasianus*, Baden Köpfl.
- BENOZZO F. (2011), *Dalla filologia tradizionale all'etnofilologia tradizionante*. In D. Fiormonte, ed., *Canoni liquidi. Variazione culturale e stabilità testuale dalla Bibbia a Internet*, ScriptaWeb, pp. 28-42.
- BODARD G., GARCÉS J. (2009). *Open Source Critical Editions: A Rationale*, in M. Deegan, K. Sutherland, edd., *Text Editing, Print, and the Digital World*, Ashgate, pp. 83-98.
- BOHNENKAMP A. ET AL. (2011). *Perspektiven auf Goethes Faust. Zur historisch-kritischen Hybridedition des Faust*, in A. Bohnenkamp, ed., *Jahrbuch des Freien Deutschen Hochstifts*, Wallstein, pp. 23-67.
- BRÜNING G., HENZEL K., PRAVIDA D. (2012). *On the dual nature of written texts and its implications for the encoding of genetic manuscripts*, a talk delivered at conference *DH 2012*. URL= <http://lecture2go.uni-hamburg.de/konferenzen/-/k/13957> [last retrieved 04.03.2013].
- BÜCHELER F., RIESE A. (1894-1926). *Anthologia latina sive poesis latinae supplementum*, B.G. Teubner.
- COZZO A. (2006). *La tribù degli antichisti: un'etnografia ad opera di un suo membro*, Carocci.
- DRISCOLL M. J. (2006). *Levels of transcription*, in L. Burnard, K. O'Brien O'Keefe, J. Unsworth, edd., *Electronic Textual Editing*, Modern Language Association of America.
- FIORMONTE D. (2012a). *Towards a Cultural Critique of Digital Humanities*. «Historical Social Research / Historische Sozial Forschung», Vol. 37 (3), n. 141 (=M. Thaller, ed., *Proceedings of conference Controversies around the Digital Humanities*), pp. 59-76. URL= http://www.cceh.uni-koeln.de/files/Fiormonte_final.pdf [last retrieved 02.03.2013].

- FIORMONTE D. (2012b). *Testo Tempo Verità*, «Humanist Studies & the Digital Age» 2 (1).
- HAUGEN O. E. (2004). *Parallel Views: Multi-level Encoding of Medieval Nordic Primary Sources*, «Literary and Linguistic Computing» 19 (1), pp. 73-91.
- LESPECT J.-F. (2005). *Vespa, le Iudicium coci et pistoris iudice Vulcano (Anthologie Latine, 199): introduction, texte latin, traduction et notes*, «Folia Electronica Classica» 9. URL= <http://bcs.fltr.ucl.ac.be/fe/09/VespaIntro.html> [last retrieved 10.03.2013].
- MORDENTI R. (2001). *Informatica e critica dei testi*, Bulzoni.
- MORDENTI R. (2011). *Paradosis. A proposito del testo informatico*, Accademia Nazionale dei Lincei.
- MORDENTI R. (2012). *Prospettive e problemi per l'edizione critica digitale dello Zibaldone Laurenziano (Plut. XXIX, 8) di Giovanni Boccaccio*, a talk delivered at the round table *L'informatica umanistica e i suoi problemi. L'edizione critica digitale dei testi. Roma, 20 giugno 2012*, Centro Linceo Interdisciplinare "B. Segre". URL= http://www.lincei.it/files/centro_linceo/Rela_Mordenti.pdf [last retrieved 05.03.2013].
- MONELLA P. (2012). *In the Tower of Babel: modelling primary sources of multi-testimonial textual transmissions*, a talk delivered at the *London Digital Classicist Seminars 2012*, Institute of Classical Studies, London, on 20.07.2012. URL= <http://www.digitalclassicist.org/wip/wip2012.html> [last retrieved 17.03.2013].
- MONELLA P. (forthcoming). *Why are there no digital scholarly editions of "classical" texts?* forthcoming in the proceedings of The fourth meeting on Digital Philology, Verona 13-15.09.2012. Pre-print: URL= <http://www.unipa.it/paolo.monella/lincei/why.html> [last retrieved 10.03.2013].
- OMONT H. A. (1903). *Anthologie de poètes latins dite de Saumaise: Reproduction réduite du manuscrit en onciale, latin 10318, de la Bibliothèque nationale*, Berthaud frères.
- ORLANDI T. (1999). *Ripartiamo dai diasistemi*, in *I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici*, *Conv. Int. 27-29 maggio 1998*, Accademia Nazionale dei Lincei, pp. 87-101.

- ORLANDI T. (2010). *Informatica testuale. Teoria e prassi*, Laterza.
- PASQUALI G. (1971). *Storia della tradizione e critica del testo*, Felice Le Monnier.
- PERRI A. (2009). *Al di là della tecnologia, la scrittura. Il caso Unicode*. «Annali dell'Università degli Studi Suor Orsola Benincasa» 2, pp. 725-748.
- PIERAZZO E. (2011). *A rationale of digital documentary editions*. «Literary and Linguistic Computing» 26 (4), pp. 463-477.
- PINI F. (1958). *Iudicium coci et pistoris*, Gismondi.
- REYNOLDS L., WILSON N. (1991). *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*, Clarendon Press.
- SAMPSON G. (1990). *Writing Systems: A Linguistic Introduction*, Stanford University Press.
- SCHMIDT D., COLOMB R. (2009). *A data structure for representing multi-version texts online*. «International Journal of Human-Computer Studies» 67 (6), pp. 497-514.
- SCHMIDT D., FIORMONTE D. (2010). *Multi-Version Documents: A Digitisation Solution For Textual Cultural Heritage Artefacts*, «Intelligenza Artificiale» 4 (1), pp. 56-61.
- SHACKLETON BAILEY D. R. (1980). *Three pieces from the Latin anthology*, «Harvard Studies in Classical Philology» 84, pp. 177-217.
- SHACKLETON BAILEY D. R. (1982). *Anthologia Latina, I: Carmina in codicibus scripta, Fasc. 1: Libri Salmasiani aliorumque carmina*, Teubner.
- THALLER M. (2006). *Strings, Texts and Meaning*, in *Digital Humanities 2006. The First ADHO International Conference*. Université Paris-Sorbonne, 5-9.07.2006. Conference Abstracts, Centre de Recherche Cultures Anglophones et Technologies de l'Information, pp. 212-214.
- WITTERN C. (2006). *Writing Systems and Character Representation*, in L. Burnard, K. O'Brien O'Keefe, J. Unsworth, edd., *Electronic Textual Editing*, Modern Language Association of America.

PARTE IV
BIBLIOTECHE DIGITALI TESTUALI:
PROGETTI E SERVIZI INNOVATIVI

11. Immaginando il futuro digitale della traduzione letteraria: la traduzione dell'edizione 1860 di *Leaves of Grass* per il Walt Whitman Archive

Caterina Bernardini¹

Abstract. Lo scopo di questo saggio è di discutere le possibilità di riconfigurazione della traduzione letteraria nell'ambito dell'umanistica digitale e di illustrare il progetto sperimentale, in via di sviluppo, di traduzione in italiano dell'edizione 1860 di *Leaves of Grass* di Walt Whitman per la pubblicazione all'interno dell'archivio letterario digitale *Walt Whitman Archive* (<http://www.whitmanarchive.org/>). A livello teorico, discuterò come l'interazione di una disciplina antica e complessa come la traduzione letteraria con le nuove risorse dell'umanistica digitale possa condurre non soltanto ad una diversa e rinnovata pratica di traduzione, ma ad un vero e proprio ripensamento della disciplina stessa e dei modi della sua rappresentazione. In questo senso, così come il concetto di 'lettura' è stato, e continua ad essere, profondamente modificato dall'umanistica digitale grazie alla teorizzazione Morettiana della 'distanza'², anche quello di 'traduzione' potrebbe essere rivisitato in termini di una 'visibilità' che risponda alla denuncia Venutiana dell'eccessiva "invisibilità" dell'istanza traduttiva³.

Parole chiave: traduzione, edizione digitale, visibilità, riconfigurazione, *Walt Whitman Archive*.

¹ Università degli Studi di Macerata, Italia. e-mail: caterina.bernardini@gmail.com
University of Nebraska-Lincoln, USA.

² Vedi Franco Moretti (2007). *Graphs, Maps, Trees. Abstract Models for a Literary History*, Verso.

³ Vedi Lawrence Venuti (1997). *The Translator's Invisibility: a History of Translation*, Routledge.

Pochi sono ancora gli interventi critici e i progetti sperimentali di umanistica digitale che si occupino di traduzione letteraria⁴, eppure molti sono i punti d'incontro, sia teorici che pratici, delle due discipline. La mia tesi di dottorato congiunto in *Modern and Comparative Languages and Literatures* (Università degli Studi di Macerata, Italia – University of Nebraska-Lincoln, USA), incentrata sullo studio della ricezione italiana dell'opera di Walt Whitman e sulla traduzione in italiano dell'edizione 1860 di *Leaves of Grass*, vuol rappresentare un primo passo in questo senso. L'edizione in questione non è mai stata pubblicata in Italia, né tradotta in italiano prima d'ora, e la scelta del contesto dell'archivio digitale per la pubblicazione mira specificamente a significare una revisione e riconfigurazione della teoria e della metodologia della traduzione e degli studi sulla ricezione in ambito digitale. Il *Walt Whitman Archive*, la cui sede fisica principale è il Center for Digital Research in the Humanities presso la University of Nebraska-Lincoln, è, insieme al *William Blake Archive* e al *Dickinson Electronic Archives* uno dei primi e, ad oggi, più completi, archivi letterari digitali americani. Fondato nel 1995 dagli attuali direttori, Prof. Kenneth Price (University of Nebraska-Lincoln) e Prof. Ed Folsom (University of Iowa), l'archivio digitale ha rivoluzionato gli studi Whitmaniani, rendendo consultabili, in *open-access*, tutte le sei edizioni di *Leaves of Grass* e dando così il meritato rilievo alla poetica di continua revisione messa in atto da Whitman e rimasta nascosta per la tendenza a studiare, tradurre e pubblicare soltanto la prima edizione del 1855 e la cosiddetta *deathbed edition* del 1891-92. L'archivio, che può contare sul sostegno di enti come il National Endowment for the Humanities e su un consorzio di Università e gruppi di ricerca che gli permettono di avere stabilmente a disposizione uno staff di più di trenta persone (tra cui studenti, ricercatori, docenti, tecnici informatici) sta gradualmente mettendo a disposizione degli utenti anche tutti gli scritti di prosa e tutta la corrispondenza Whitmaniana, nonché i manoscritti delle bozze relative al processo di revisione del poeta di Camden. Come osservato dal Professor Ken Price, l'archivio elettronico in continua espansione si presta particolarmente a rappresentare

⁴ Si pensi ad esempio al progetto *Digital Dante* della Columbia University (<http://dante.ilt.columbia.edu/comedy/index.html>), che riporta varie traduzioni della *Divina Commedia* in inglese e permette di fare confronti diretti delle differenti rese traduttive.

l'opera Whitmaniana, dal momento che "it defies the constraints of the book. Whitman's work was always being revised, was always in flux, and fixed forms of print do not adequately capture his incessant revision" ("About the Archive" in <http://www.whitmanarchive.org/about/history.html>).

Si sta anche via via procedendo ad espandere la sezione dedicata alle traduzioni di *Leaves of Grass* in varie lingue straniere, ed è stato aperto uno spazio sperimentale dedicato allo studio comparato della ricezione e traduzione di uno stesso testo Whitmaniano (la prima poesia presa in esame è "Poets to Come") in più lingue (in <http://whitmanarchive.org/published/foreign/index.html>).

L'esperimento ha messo in evidenza l'interdipendenza delle varie traduzioni (per esempio, la traduzione in lingua spagnola di Vasseur sembra profondamente influenzata da quella italiana di Gamberale), come anche l'intrecciarsi, spesso inaspettato, delle varie ricezioni 'nazionali' nella fluidità di una comune mediazione intellettuale.

L'archivio si va sempre di più configurando, dunque, non solo come una *repository*, non solo come una collezione archivistica dei materiali Whitmaniani, ma come un vero e proprio esperimento euristico collettivo volto a cambiare, innovare, ridefinire, il *modus operandi* della critica Whitmaniana passata, presente e futura. Il mio progetto di traduzione parte, in questo senso, dall'assunto secondo cui l'umanistica digitale non è umanistica trasferita nel digitale, non è umanistica che del digitale usa i nuovi strumenti tecnologici per ottenere vecchi risultati. Essa è, invece, un'umanistica che risulta profondamente rinnovata e trasformata dall'interazione col digitale, non solo nei risultati a cui giunge, ma anche nei suoi fondamenti. Come scrivono Anne Burdick *et al.*, discutendo le prospettive della seconda generazione (quella in atto) di umanisti digitali,

Maintaining criticality and experimentation means challenging received traditions, even – perhaps, especially – those that defined the first generation of Digital Humanities Work. Innovative forms of public engagement, new publishing models, imaginative ways of structuring humanistic work, and new units of argument will come to take their place beside the pioneering projects of the first generation. This means embracing new skill-sets that are not necessarily associated with traditional humanistic training: design, programming, statistical

analysis, data visualization, and data-mining. And this means developing new humanities-specific ways of modeling knowledge and interpretation in the digital domain. It means showing that interpretation is rethought through the encounter with computational methods and that computational methods are rethought through the encounter with humanistic modes of knowing. (2012, 103).

Il digitale offre infinite possibilità alla traduzione, ma non si tratta di un processo di facilitazione di pratiche traduttive e strategie editoriali che finiscono per rimanere essenzialmente tradizionali. Si tratta piuttosto di un ripensamento vero e proprio della traduzione, di una sua ridefinizione come disciplina caratterizzata da un continuo passare attraverso processi di trasformazione, come scriveva Walter Benjamin nel saggio intitolato "On Language as Such". In questo modo si realizzerebbe il pronostico della studiosa di traduzione Emily Apter, secondo cui "Digital technology, it would seem, is increasingly challenging the boundaries of what translation is" (2005, 10). Del resto, l'interazione dell'umanistica digitale con la traduzione letteraria sembra auspicabile anche per altri motivi. Le due aree coincidono, ad esempio, per esser caratterizzate da una certa ansietà di raggiungere una definizione teorica adeguata alla propria natura disciplinare, ma anche per la tendenza a produrre interpretazioni e argomentazioni 'concrete' piuttosto che teoriche. Inoltre, vi è in entrambe un certo entusiasmo utopico circa le proprie potenzialità, e un corrispondente senso distopico dei limiti di quello che si può e non si può fare, considerate le proprie risorse e la propria 'lingua': una lingua che deve sempre essere rispettata, ma che allo stesso tempo può essere sfidata e rinnovata. Ma chi deve bussare alla porta di chi? Dal momento che ad oggi sono pochi i traduttori che siano anche attivamente impegnati nell'umanistica digitale, e viceversa, come può avvenire, quest'interazione? A me sembra che siano i traduttori a doversi muovere in questo senso. Come suggerito dalla studiosa di teoria culturale, nonché *media designer* Anne Balsamo nel suo *Designing Culture: the Technological Imagination at Work*, spetta agli umanisti 'tradizionali' progettare nuovi strumenti tecnologici per il proprio lavoro e la propria ricerca, non rimanendo ad aspettare strumenti che essi potrebbero poi giudicare inadeguati o non pertinenti, e impegnandosi invece attivamente nell'immaginarne di nuovi. Allo stesso tempo, gli

umanisti digitali possono guardare alla traduzione come un territorio promettente nel quale sperimentare e costruire nuove e rivoluzionarie tecnologie.

Come osserva Matthew Kirschenbaum, "Digital Humanities is about a scholarship and a pedagogy that are publicly visible in ways to which we are generally unaccustomed" (2012, 4). Questo è un concetto affascinante, se lo si applica alla traduzione letteraria. Secondo lo studioso di traduzione Lawrence Venuti, l'invisibilità tradizionalmente imposta al traduttore e ai suoi processi decisionali, nonché gli ingenui ideali di trasparenza e fedeltà al testo originale, hanno contribuito a relegare la traduzione letteraria in una posizione di secondo grado rispetto a molte altre discipline umanistiche, come se fosse una semplice attività meccanica di servizio e di trasposizione comunicativa derivata e neutrale. Scrive Venuti:

"Invisibility" is the term I will use to describe the translator's situation and activity in contemporary Anglo-American culture. It refers to two mutually determining phenomena: one is an illusionistic effect of discourse, of the translator's own manipulation of English; the other is the practice of reading and evaluating translations [...] A translated text, whether prose or poetry, fiction or nonfiction, is judged acceptable by most publishers, reviewers, and readers when it reads fluently, when the absence of any linguistic or stylistic peculiarities makes it seem transparent, giving the appearance that it reflects the foreign writer's personality or intention or the essential meaning of the foreign text – the appearance, in other words, that the translation is not in fact a translation, but the "original". The illusion of transparency is an effect of fluent discourse, of the translator's effort to insure easy readability by adhering to current usage, maintaining continuous syntax, fixing a precise meaning. What is so remarkable here is that this illusory effect conceals the numerous conditions under which the translation is made, starting with the translator's crucial intervention in the foreign text. The more fluent the translation, the more invisible the translator, and, presumably, the more visible the writer or meaning of the foreign text. (1997, 1)

La recente comparsa dello strumento di traduzione automatica chiamato Google Translate e di altri servizi simili, seppur indubbiamente utili per traduzioni veloci e, appunto, 'di servizio', non aiuta, in

questo senso, a rappresentare e comprendere la complessità della traduzione di testi letterari. La traduzione letteraria ha bisogno, come sottolineato da Venuti, di essere resa visibile: i suoi processi interpretativi e decisionali devono essere resi noti ed esibiti, in modo da demistificare l'idea classica di traduzione come operazione finita, neutrale, autonoma. Ciò si accorda chiaramente con le idee della umanista digitale Johanna Drucker, secondo la quale

One of the major aims of speculative computing and digital scholarly editions is to redesign representations of subjectivity within knowledge production. Making visible the subjective acts of interpretation and the role of imaginative play means to challenge the authority claims of formal logical systems, so that DH can extend the theoretical questions that came into focus in deconstruction, postmodern theory, critical and cultural studies, and other theoretical inquiries (2009, XIV).

La traduzione deve svelare la sua non-finitezza, la sua non-autonomia, per esempio offrendo, nel contesto digitale, apparati critici in continuo aggiornamento che le edizioni cartacee non sono in grado di offrire. Gli apparati critici possono contenere delle sezioni dedicate alle soluzioni alternative, digressioni analitiche ed esplicative, sezioni contestuali, campi semantici. Inoltre, la (iper)testualità digitale si addice alla traduzione letteraria per la possibilità di effettuare continue revisioni e modifiche, anche se vi è un rischio ipertrofico in continuo agguato. Ci si può aprire all'interazione con i lettori, inserire, come nel micro-esempio di "Poets to Come", la propria traduzione in una rete di comparazione con altre traduzioni dello stesso testo, costruire, intorno alla traduzione, una cornice storico-letteraria relativa al contesto della ricezione di una certa opera.

Certo, ciò non vuol dire che la traduzione digitale deve diventare un campo minato di spiegazioni, alternative, digressioni, commentari eccessivamente zelanti, né dovrebbe rischiare di rappresentare un'occasione catartica per una sorta di auto-commiserazione dei traduttori circa le difficoltà di un compito tanto difficile come quello del tradurre. Può essere, però, un'opportunità sperimentale per provare a mettere in pratica la missione di far corrispondere la traduzione ad una "zone of critical engagement that connects the 'I' and the 'n' of transLation and transNational" (Apter 2005, 5) e, in questo senso, all'occasione di innesco di un vero e proprio ripensamento delle lingue, letterature e

culture nel contesto della *technological literacy* o *electracy*, per dirla con Gregory Ulmer.

Ma volendo andare oltre tali attraenti premesse generali circa le potenzialità della riconfigurazione della traduzione letteraria in ambito digitale, si può tracciare un profilo metodologico di quest'ultima? Si può proporre, insomma, un'agenda della traduzione letteraria digitale? Sì e no. Si può invitare, appunto, ad aumentare la visibilità dell'istanza traduttiva, al riflettere metatraduttivo sulla teleologia dell'atto di traduzione, e all'inserimento della traduzione in un *network* ipertestuale e intertestuale che renda atto della complessità dell'atto traduttivo e del suo carattere di mediazione non solo linguistica, ma più largamente culturale. La creazione e configurazione di strumenti e tecnologie *ad hoc*, però, rimane strettamente dipendente dai progetti specifici. Come osservava il poeta e traduttore Mario Luzi, la traduzione letteraria resiste ogni sistematizzazione tecnica e metodologica: la traduzione 'avviene' quasi per miracolo, attraverso lunghi e complessi processi decisionali basati sul binomio dialettico dell'identità e della differenza. La traduzione è sempre da fare e da rifare, cercando di toccare quel punto di contatto con l'Altro che Walter Benjamin descrive in questi termini:

Just as a tangent touches a circle fleetingly and at only a single point, and just as this contact, not the point, prescribes the law in accord with which the tangent pursues its path into the infinite, in the same way a translation touches the original fleetingly and only at the infinitely small point of sense, in order to follow its path in accord with the law of fidelity in the freedom of linguistic development (*The Task of the Translator*, in Venuti, 2012, 82).

Esibire il fallimento intrinseco a questa missione di mediazione persa in partenza, ma sempre perfettibile, dovrebbe essere uno degli scopi primari della traduzione letteraria digitale, insieme al redimere il proprio *status* di attività interpretativa fondamentale e irrinunciabile per le storie delle letterature, per la letteratura comparata e per le storie delle ricezioni. Quel che è certo è che la traduzione letteraria digitale dev'essere prima di tutto immaginata, *disegnata*, poiché, come osserva l'umanista digitale Christine L. Borgman (e ciò vale di certo anche nel nostro caso), "design decisions made today will determine whether the Internet of tomorrow enables imaginative new forms of scholarship and

learning – or whether it simply reinforces today's tasks, practices, laws, business models, and incentives" (2007, 3).

Il progetto di pubblicazione della traduzione dell'edizione 1860 sul *Walt Whitman Archive* prevede la collaborazione di traduttori, informatici, italianisti e americanisti. L'inserimento della traduzione in linguaggio xslt permetterà di realizzare un ricco *encoding* di *metadata*, e particolare rilievo verrà dato all'inserimento di collegamenti ipertestuali con altre versioni italiane dell'opera di Walt Whitman (dando così rilievo, anche per il pubblico italiano, all'esistenza delle sei edizioni delle *Leaves*) ma anche con altre traduzioni in lingue diverse, e, a livello intertestuale, con testi letterari di altri autori, italiani e stranieri, che si rivelino affini ai testi in questione, o comunque rilevanti a livello comparativo. Grafici complessi verranno inseriti col fine di mostrare le soluzioni traduttive alternative, di nuovo, sia italiane, che straniere. Quest'ultima parte mira a corrispondere a un'apertura metodologica innovativa per lo studio comparato delle traduzioni: un ambito di ricerca accademica ancora più marginalizzato di quelli della teoria della traduzione o della pratica della traduzione. Una corposa rivalutazione di tale ambito svelerebbe senza dubbio l'esistenza di molti sentieri mai solcati prima d'ora dalla critica letteraria e dalla letteratura comparata. Uno degli elementi centrali per la traduzione letteraria, ovvero il ritmo ⁵, la trasformazione fonotestuale di un testo di partenza in uno di arrivo che la traduzione mette in atto, potrà essere studiato e divenire un importante termine di comparazione nello studio del rapporto tra testo di partenza e testo di arrivo e di raffronto di traduzioni diverse di uno stesso testo, grazie a nuove strumentazioni per l'analisi del suono e degli schemi ritmici, come quelle che l'umanista digitale Tanya Clement sta sperimentando all'interno del progetto HIPSTAS (<http://blogs.ischool.utexas.edu/hipstas/2012/11/14/welcome-to-hipstas/>). Lo studio delle ricorrenze ritmiche aprirà senza dubbio nuovi orizzonti non solo per la teoria e prassi della traduzione, ma anche per la critica letteraria, più in generale.

Ma volendo giungere ad un esempio concreto, prendiamo il caso di uno dei versi Whitmaniani più problematici per i traduttori di tutto il mondo: "I sound my barbaric yawp over the roofs of the world" ("Song

⁵ Si pensi, ad esempio, alla centralità assegnata all'elemento ritmico nella poetica traduttiva di Henri Meschonnic in *Poétique du Traduire*, Verdier, 1999.

of Myself" ...) e immaginiamone una traduzione digitale 'visibile'. "Yawp" è un termine di natura fortemente onomatopeica, e piuttosto inusuale per l'inglese americano di Whitman nonché per quello a noi contemporaneo. Esso indica una specie di ululato, quasi di natura animalesca, quasi un latrato. L'italiano non presenta termini 'direttamente' equivalenti, e i traduttori italiani (il verso in questione, infatti, ricorre anche in tutte le edizioni delle *Leaves of Grass* finora tradotte in italiano) hanno optato per generici urli, grida, strilli. Le varie soluzioni adottate nelle traduzioni passate potrebbero essere offerte all'interno di un grafico che potrebbe anche indicare, seppur schematicamente, le interconnessioni tra le varie soluzioni. Un link ipertestuale potrebbe collegare la parola del testo originale alla sua definizione nel dizionario Webster usato da Walt Whitman, ma anche ad eventuali ricorrenze di termini nei giornali dell'epoca ed in altri testi letterari, mentre una nota, contenente un altro link, potrebbe rimandare all'influenza di questo verso per la poetica di Allen Ginsberg, e, in particolare, per il suo "Howl", ma anche alla scena del film di Peter Weir "Dead Poets Society" in cui si cita proprio questo verso. La traduzione da me proposta, "Io lancia il mio *ioup* barbarico sopra i tetti del mondo" sarebbe corredata di un commento che spiegherebbe le ragioni della scelta di *ioup*, fatta, sostanzialmente, per privilegiare l'effetto fonico straniante e non privare i lettori italiani della possibilità di pronunciare il termine, che altrimenti, nella versione originale, risulterebbe inaccessibile per la gran parte di essi. Inoltre, la possibilità dell'apertura di ulteriori finestre potrebbe permettere ai lettori di approfondire aspetti relativi alla ricezione italiana di questi versi e di aprire collegamenti ipertestuali riguardanti, ad esempio, la ricorrenza dell'aggettivo "barbaro" o "barbarico" nella letteratura italiana di fine Ottocento e inizio Novecento, o la ricorrenza di questo verso nella critica italiana su Walt Whitman. Chiaramente, come ho già detto, l'obiettivo non è quello di offrire ai lettori un testo pesantemente minato da interventi critici ed editoriali: il solo testo tradotto in modo "classico" rimarrebbe consultabile, ma allo stesso tempo i lettori sarebbero consapevoli del sostrato culturale, interpretativo, decisionale, di ogni singola parola e di ogni singolo verso, e potrebbero ogni volta scegliere di aprire, o non aprire, le finestre ad essi collegate. In questo senso, la traduzione si farebbe più 'visibile'. E chiaramente, sarebbe possibile arricchirla e riconfigurarla

costantemente, e i lettori stessi potrebbero apportare materiale, commenti, osservazioni, intuizioni.

Visione utopica, perfezionismo, pedanteria? Forse. Quel che è certo è che la traduzione letteraria, marginalizzata dall'accademia, strumentalizzata dall'editoria e perennemente mal rappresentata, ha bisogno di una riconfigurazione teorica e pratica, e che l'umanistica digitale sembra, in questo senso, un'occasione da non perdere. E sebbene sia necessaria prima di tutto l'immaginazione, per progettare nuovi strumenti e sistemi che aiutino a trasformare quest'opportunità in successo, solo un contatto profondo e autentico dei traduttori con l'umanistica digitale farà sì che quest'interazione assuma davvero una portata rivoluzionaria.

11.1. Bibliografia

11.1.1. Monografie

- APTER E. (2006). *The Translation Zone: A New Comparative Literature*, Princeton University Press.
- BALSAMO A. (2011). *Designing Culture: The Technological Imagination at Work*, Duke University Press.
- BARTSCHERER T. & RODERICK COOVER (2011). *Switching Codes: Thinking Through Digital Technology and the Arts*, The University of Chicago Press.
- BENJAMIN W. (1997). *One way Street and Other Writings*, Verso.
- BORGMAN C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*, The MIT Press.
- BURDICK A. ET AL. (2012). *Digital Humanities*, The MIT Press.
- CONDON F. ET AL. (2001). *Digital Resources for the Humanities*, West Virginia University Press.
- DRUCKER J. (2009). *Speclab. Digital Aesthetics and Projects in Speculative Computing*, Chicago University Press.
- EARHART A. & A. JEWELL (EDS) (2011). *The American Literature Scholar in the Digital Age*, The University of Michigan Press.
- HAYLES, K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*, University of Chicago Press.

- HOCKEY S. (2006). *Electronic Texts in the Humanities*, Oxford UP.
- MESCHONNIC H. (1999). *Poétique du traduire*, Verdier.
- MORETTI F. (2007). *Graphs, Maps, Trees. Abstract Models for a Literary History*, Verso.
- VENUTI L. (1997). *The Translator's Invisibility: A History of Translation*, Routledge.
- _____ (2012). *The Translation Studies Reader*, Routledge.
- SCHRIEBMAN S. & RAY SIEMENS (EDS) (2007). *A Companion to Digital Literary Studies Humanities*, Blackwell.
- ULMER G. L. (2003). *Internet Invention From Literacy to Electracy*, Longman.

11.1.2. Contributi in volume

- KIRSCHENBAUM M. (2012). *What Is Digital Humanities and What It's Doing in English Departments?* In Gold M. K. (ed) *Debates in the Digital Humanities*, University of Minnesota Press, pp. 3-11.
- LUZI, M. (2004). *Riflessioni sulla traduzione*. In Buffoni F. *La traduzione del testo poetico*. Marcos y Marcos, pp. 48-54.

12. La localizzazione del Perseus Project in lingua italiana

Federico Boschetti¹

Abstract. Si illustra il progetto di localizzazione in lingua italiana dell'infrastruttura per lo studio dei classici greci e latini costituita dal Perseus Project (Tufts University, Medford, MA), usando risorse per l'analisi della nostra lingua sviluppate presso l'ILC-CNR di Pisa e mettendo a disposizione nuovi componenti software per la visualizzazione e l'annotazione di testi bilingui.

Parole chiave: filologia computazionale, allineamento, lessico dinamico.

12.1. Introduzione

La localizzazione in lingua italiana del Perseus Project² è l'obiettivo di un progetto congiunto che coinvolge la Tufts University di Medford, MA e l'Istituto di Linguistica Computazionale "A. Zampolli" del CNR di Pisa³.

Le diverse fasi del progetto prevedono a) l'acquisizione delle risorse, b) l'analisi linguistica dei documenti in lingua italiana, c) l'individuazione delle corrispondenze fra le due lingue in oggetto a diversi livelli di granularità, d) la creazione di lessici bilingui dinamici, e) la visualizzazione delle traduzioni messe in parallelo con gli originali e

¹ ILC/CNR, Pisa, Italia. e-mail: federico@ilc.cnr.it.

² Per una panoramica sulle attività del Perseus Project si veda Babeu (2011). Il Perseus Project è accessibile in rete all'indirizzo <http://www.perseus.tufts.edu> [ultima visita 1.3.2013].

³ Il progetto è finanziato dal National Endowment for the Humanities (NEH) e dal Consiglio Nazionale delle Ricerche (CNR).

f) la possibilità di annotare i testi da parte di studiosi e studenti.

I principali obiettivi di ricerca consistono nell'affinamento delle tecniche per l'allineare gli originali alle traduzioni plurilingui, la creazione di lessici dinamici e la comparazione di analisi linguistiche e stilistiche nelle lingue originali e nelle traduzioni. Ma gli scopi di ricerca propri della filologia computazionale si coniugano anche con gli obiettivi didattici che caratterizzano i paradigmi della *peer education*, in quanto vengono messi a disposizione strumenti collaborativi per l'apprendimento online, promuovendo la partecipazione attiva necessaria alla creazione di nuovi contenuti. In primo luogo, si mira a rendere più accessibili agli studenti italiani del Liceo e dei primi anni dell'Università gli strumenti digitali di analisi e interpretazione dei testi antichi già messi a disposizione dal Perseus Project in lingua originale e in traduzione inglese. In secondo luogo, si intende coinvolgere gli studenti nel processo di produzione e validazione di contenuti relativi agli studi classici, con la possibilità di tener traccia (col dovuto rispetto della *privacy*) dei contributi individuali, anche in vista di un futuro *e-portfolio*⁴ che possa essere riconosciuto internazionalmente, costituendo titolo di merito in particolare per soggiorni di formazione all'estero.

12.2. I repertori di testi

Il progetto si basa su materiali disponibili online, con licenza aperta, in particolare Creative Commons⁵ - Attribution - Noncommercial - Share - Alike (CC BY-NC-SA 3.0). I repertori di testi cui si fa ricorso in modo sistematico per acquisire traduzioni di classici non più coperte da copyright sono di vario tipo per accuratezza, estensione e metodi di lavoro.

La Biblioteca Italiana⁶ mette a disposizione testi digitali, con marcatura T.E.I., di ottima qualità ma, coerentemente con i suoi scopi, vi si trovano soltanto traduzioni che siano rilevanti per la letteratura Italiana,

⁴ Si veda, per un approfondimento sul concetto di *e-portfolio*, Crane (2012).

⁵ Si veda <http://creativecommons.org> [ultima visita 27.2.2013] per una panoramica completa sulle licenze Creative Commons.

⁶ <http://www.bibliotecaitaliana.it> [ultima visita 27.2.2013]

come la traduzione dell'Iliade di Monti. LiberLiber⁷ mette a disposizione un numero elevato di traduzioni digitalizzate, con diversi livelli di accuratezza rispetto agli originali cartacei, dichiarati nei metadati. Filosofico.net⁸ mette a disposizione traduzioni in Italiano di testi prevalentemente filosofici, ma in alcuni casi è possibile trovare traduzioni di tragedie, come l'*Elena* di Euripide, non presenti in altri repertori. Il sito è interessato a diffondere contenuti, con minore attenzione al processo di digitalizzazione, revisione e scrittura di metadati riguardanti le opere messe a disposizione con licenza libera. Un discorso a parte merita la sezione italiana di Wiki-Source, di cui si parla per esteso nella prossima sezione.

12.3. Il contributo di WikiSource

WikiSource⁹ è una comunità di proof-reading suddivisa in sottocomunità nazionali che si occupano della digitalizzazione e della revisione di testi scritte nelle diverse lingue. WikiSource mette a disposizione dei propri membri il testo acquisito tramite OCR affiancato all'immagine della pagina dell'edizione cartacea, in modo tale che la conformità della copia digitale all'originale possa essere sottoposto a continua verifica e miglioramento. Ciascuna pagina è trascritta, formattata e riletta, e lo stato di avanzamento del lavoro viene registrato e segnalato. La validazione finale non può essere fatta dallo stesso utente che ha compiuto le precedenti operazioni, per dare garanzia di un reale processo di revisione che coinvolga almeno due responsabili.

Le scansioni sono scaricate dai maggiori archivi online, quali Internet Archive¹⁰, Hathi Trust¹¹ e Google Books¹², oppure da OPAL¹³, portale della biblioteca dell'Università di Torino, molto attivo per la scansione di testi classici, che riversa periodicamente i propri materiali

⁷ <http://www.liberliber.it> [ultima visita 27.2.2013]

⁸ <http://www.filosofico.net> [ultima visita 27.2.2013]

⁹ <http://it.wikisource.org> [ultima visita 27.2.2013]

¹⁰ <http://archive.org> [ultima visita 27.2.2013]

¹¹ <http://hathitrust.org> [ultima visita 27.2.2013]

¹² <http://books.google.com> [ultima visita 27.2.2013]

¹³ <http://www.opal.unito.it> [ultima visita 27.2.2013]

in archivi general purpose, come Internet Archive. L'OCR, quando il testo digitalizzato non sia già disponibile o non sia soddisfacente, viene eseguito con software commerciale (ad es. Fine Reader) o, preferibilmente, con software libero (ad es. Tesseract¹⁴).

WikiSource Italia sta offrendo la propria collaborazione esterna al progetto, grazie alla correzione dell'OCR di traduzioni italiane di classici greci e latini.

12.4. Le traduzioni

Il Perseus Project mette a disposizione di studiosi e studenti, oltre ai testi classici in lingua originale, anche traduzioni in lingua inglese non più coperte da copyright. Le edizioni più recenti, quindi, non sono mai posteriori alla metà del secolo scorso. Anche le traduzioni in lingua italiana che saranno impiegate nel presente progetto sono principalmente da collocare fra la fine dell'Ottocento e i primi decenni del Novecento, ad esclusione di traduzioni letterarie di particolare rilievo, come ad esempio la traduzione dell'*Eneide* di Annibal Caro.

Un rapido confronto tramite indice Gulpease e indici di frequenza fra l'*Agamennone* di Eschilo tradotto in lingua italiana da Romagnoli (1922) e la medesima opera tradotta in lingua inglese da Smyth (1926), pur senza alcuna pretesa di rilevanza statistica, dato che studi più accurati saranno svolti in fase più avanzata del lavoro, rafforza tuttavia l'impressione che le traduzioni italiane siano molto più lontane dalla lingua d'uso rispetto alle traduzioni inglesi. Qui si intende dare solo alcuni icastici esempi.

Nel prologo del suo *Agamennone* (vv. 1-27), Romagnoli usa forme quali *gomita* (=gomiti), *veggo* (=vedo), *riscintillanti* (=scintillanti), *ètra* (=cielo), *che ... fulgono* (=che ... risplendono), *verno* (=inverno). Smyth, nella sua traduzione in prosa, usa rispettivamente *arm*, *I have learned to know*, *radiant*, *conspicuous*, *firmament*, *winter*.

Di segno opposto l'operazione di Acri (1913-15), che infarcisce la sua traduzione dei *Dialoghi* platonici di toscanismi, per rendere (maldestramente per l'orecchio contemporaneo) l'espressività del parlato. Alla fine dell'*Eutifrone* (15e) si assiste a questo scambio di battute tra i due ateniesi:

¹⁴ <https://code.google.com/p/tesseract-ocr> [ultima visita 28.2.2013]

SOCRATE [...] Deh, parla, bravo Eutifrone, e non mi tener piú nascosto ciò che ne pensi.

EUTIFRONE A un'altra volta, Socrate: ho furia; l'è ora ch'io vada. Linguisticamente neutra, priva di connotazioni regionali e vicina all'uso corrente dei parlanti lingua inglese è invece la traduzione di North Fowler (1913):

SOCRATES [...] So tell me, most excellent Euthyphro, and do not conceal your thought.

EUTHYPHRO Some other time, Socrates. Now I am in a hurry and it is time for me to go.

Date le finalità didattiche (oltre che di ricerca), è necessario capire quanto sia effettivamente proficuo mettere a disposizione degli studenti italiani, a causa di vincoli sul copyright, traduzioni che spingono ad allontanarsi piuttosto che ad avvicinarsi al testo. Per questo motivo si sta cercando di trasformare una potenziale barriera in una risorsa, tramite la collaborazione con docenti dei licei¹⁵ che stimolino gli allievi a lavorare sull'originale greco e sulla traduzione italiana congiuntamente presi, usando tutti gli strumenti di analisi messi a disposizione dal Perseus Project (analisi morfologica o morfosintattica (quando presente), Liddell-Scott James online, traduzioni in Inglese) per annotare sia il Greco che l'Italiano. Nello spirito dell'*e-portfolio* di cui si parlava nell'introduzione, gli studenti hanno l'opportunità di collaborare, con la supervisione dell'insegnante e in modo adeguato al proprio livello di formazione, a nuove traduzioni dei testi che stanno studiando, che andranno ad arricchire, dopo opportuna revisione, i repertori di traduzioni messe a disposizione con licenza CC BY-NC-SA 3.0.

12.5. L'allineamento a granularità più grezza

L'allineamento del testo in lingua originale con la traduzione italiana viene effettuato a diversi livelli di granularità, per differenti destinazioni d'uso. Nella prima fase, attualmente in corso, i testi vengono suddivisi in pericopi di lunghezza variabile, da un singolo periodo sintattico ad un intero paragrafo. Euristiche per la suddivi-

¹⁵ I primi esperimenti si stanno avviando con il Liceo Prati di Trento sull'*Elena* di Euripide.

sione automatica sono applicate ad opere che presentino, tanto nell'originale quanto nella traduzione, suddivisioni logiche organizzate gerarchicamente (libro - parte - capitolo - sezione - sottosezione - paragrafo) oppure strutture dialogiche, con eventuali ulteriori suddivisioni in paragrafi (per testi drammatici o dialoghi platonici).

Dato che la traduzione italiana in molti casi non è basata sull'edizione critica in lingua originale messa a disposizione dal Perseus Project, l'allineamento delle pericopi è affidato ad una applicazione che usa i consueti algoritmi sviluppati in bioinformatica per l'allineamento di sequenze di DNA¹⁶. Diversa attribuzione di alcune battute, traduzione di sezioni considerate spurie e omesse nell'edizione da allineare o altri interventi editoriali rendono necessari, infatti, aggiustamenti nelle due sequenze che si devono corrispondere.

La somiglianza delle pericopi nell'una e nell'altra lingua è valutata in modo molto semplice ma sufficientemente efficace tramite due criteri: somiglianza di un elemento paratestuale e somiglianza proporzionale dell'estensione delle pericopi. La numerazione di odi, sezioni, paragrafi, tende a corrispondersi con differenza minima, quando non sia nulla, a causa di slittamenti nella progressione per inserzione od omissione di unità. Si corrispondono inoltre, in testi drammatici (o dialogici), i nomi dei parlanti, tradotti tramite semplici indici bilingui delle *dramatis personae*. Soddisfatto il primo criterio, il secondo criterio si basa sull'ipotesi che pericopi di una certa lunghezza in una lingua corrispondano a pericopi di lunghezza proporzionalmente simile nella lingua che la traduce: ci si aspetta che un lungo discorso di Socrate che occupa il cinque per cento di un dialogo platonico in lingua originale occupi non molto di più e non molto meno del cinque per cento della traduzione.

Il testo della pericope in lingua originale e della pericope che la traduce viene lemmatizzato rispettivamente con gli strumenti d'analisi morfologica sviluppati presso il Perseus Project e con gli strumenti di analisi dell'Italiano sviluppati presso l'ILC, per i quali si veda Battista - Pirrelli (1999), Dell'Orletta *et al.* (2007), Dell'Orletta (2009). Le pericopi sono quindi caricate su *Aporia*, la *web application* per la visualizzazione, ricerca e annotazione di testi paralleli sviluppata originariamente dal gruppo di filologia computazionale dell'ILC per l'allinea-

¹⁶ La libreria creata per l'allineamento di oggetti generici, tramite la valutazione della somiglianza delle caratteristiche di tali oggetti, è disponibile online all'indirizzo <http://github.com/CoPhi/aligner>

mento di testi greci ed arabi.

La Fig. 12.1 illustra un esempio di pericopi allineate, mostrando come le porzioni di testo possano essere variabili e come ad una delle due pericopi possa non corrispondere testo nell'altro (nel caso specifico per il fatto che la traduzione introduce una didascalia scenica).

<p>Φύλαξ θεοὺς μὲν αἰῶ τῶνδ' ἀπαλλαγὴν πόνων φρουρᾶς ἑτέρας μήκος, ἦν κοιώμενος στέγαις Ατρείδων ἄγκαθεν, κυνὸς δίκην, ἄσπρων κάτωδα νυκτέρων ὀμήγγυριν, καὶ τοὺς φέροντας χεῖμα καὶ θέρος βροτοῖς 5 λαμπροὺς δυνάστας, ἐμπρέποντας αἰθέρι ἀστέρας, ὅταν φθίνωσιν, ἀντολάς τε τῶν.</p>	<p>SCOLTA: Numi, il riscatto concedete a me dei miei travagli, della guardia lunga un anno già, ch'io vigilo sui tetti degli Atridi, prostrato su le gomita a mo' d'un cane. E de le stelle veggio 5 il notturno concilio, ed i signori riscintillanti che nell'etra fulgono, ed il verno e la state all'uorte recano.</p>
<p>καὶ νῦν φυλάσσω λαμπάδος τὸ σύμβολον, αὐγὴν πυρὸς φέρουσαν ἐκ Τροίας φάτιν ἀλώσιμόν τε βᾶξιν· ὦδε γὰρ κρατῆ 10 γυναικὸς ἀνδρόβουλον ἐλπίζον κέαρ. εὖτ' ἂν δὲ νυκτίπλαγκτον ἔνδροσόν τ' ἔχω εὐνήν ὄνειροις οὐκ ἐπισκοπούμενην ἐμὴν· φόβος γὰρ ἀνθ' ἕπνου παραστατῆ, τὸ μὴ βεβαίως βλέφαρα συμβαλεῖν ὑπνω· 15 ὅταν δ' ἀείδῃν ἢ μινύρεσθαι δοκῶ, ἕπνου τὸ δ' ἀντιμόλπον ἐντέμων ἄκος, κλαίω τὸτ' οἴκου τοῦδε συμφορὰν στένων οὐχ ὡς τὰ πρόσθ' ἄριστα διαπονουμένου. νῦν δ' εὐτυχὴς γένοιτ' ἀπαλλαγὴ πόνων 20 εὐαγγέλου φανέντος ὀρφναίου πυρὸς.</p>	<p>Ed ora il segno aspetto della lampada, del fuoco il raggio, che da Troia rechi 10 della presa città la fama e il grido. Così comanda il cuor che aspetta e brama di maschia donna. E intanto, ecco il mio letto, irrequieto, molle di rugiada, né sogno alcuno lo frequenta mai: 15 ché non sovrasta a me sonno, ma tema ch'io le pupille a sopor greve chiuda. E quando intono - a cogliere un antidoto che il sonno vinca - un canto od una nenia, io gemo allora, e piango la ventura 20 di questa casa, che non è più retta, come già fu, pel meglio. Ed ora giunga, giunga felice dei travagli il termine, col fausto annunzio del notturno fuoco.</p>
	<p>(Lunga pausa. Poi, sulla cima del colle Aracneo, che incombe sulla città, s'accende e giganteggia un'immensa fiammata)</p>
<p>ὦ χαῖρε λαμπτήρ νυκτός, ἡμερήσιον φῶς πιφάυσκων καὶ χορῶν κατάστασιν πολλῶν ἐν Ἄργει, τῆσδε συμφορᾶς χάριν. ιοὺ ἰοὺ.</p>	<p>Oh! Salve, fiamma, che diurna luce 25 annunzi nella notte, e danze in Argo, danze, mercè di questa sorte fausta! Evviva! Evviva!</p>

Fig. 12.1. Visualizzazione dei testi paralleli

12.6. Annotazione

I testi in parallelo sono messi a disposizione di studiosi e studenti per la ricerca e l'annotazione. Come si vede in Fig. 12.2, è possibile selezionare sequenze di caratteri nell'una, nell'altra lingua o in entrambe e creare un commento ad esse riferito. La granularità della selezione cui l'annotazione è associata va dal singolo carattere alla coppia di intere pericopi, consentendo flessibilità di utilizzo. I metadati relativi al nome dell'u-tente, orario della creazione o modifica e numero di versione dell'anno-tazione sono salvati dal sistema, permettendo di operare individualmente o in gruppi di lavoro fra loro indipendenti. Come si vede in Fig. 12.3, ciascuna

annotazione può essere classificata con *tag* definiti dall'u-tente che permettono la creazione di indici analitici e tematici non solo per la successiva ricerca da parte degli utenti dei materiali annotati, ma anche e soprattutto per acquisire informazioni strutturate e semi-strutturate da gestire con sistemi automatici.

Lo strumento è quindi sufficientemente versatile per molteplici impieghi, anche non previsti *a priori*, sia di critica del testo che di attività esegetica e la classificazione delle annotazioni permette di operare in modo asincrono su più aspetti di queste attività. Ecco allora che un singolo studioso può registrare varianti testuali¹⁷, un gruppo di ricerca può occuparsi del commento critico del testo e una classe di studenti può contribuire, più modestamente, all'individuazione di alcuni tipi di figure retoriche nell'originale e nella traduzione, per metterle a confronto e ragionare sulla resa più o meno efficace.

In Fig. 12.3 è possibile inoltre vedere che le parole non riconosciute dallo spell-checker sono evidenziate. Usando dizionari-utente preparati per rilevare specifici fenomeni linguistici, l'attenzione degli studenti è diretta verso tali fenomeni da commentare (in questo caso, forme poetiche lontani dalla lingua d'uso).

The screenshot shows a web-based annotation tool. On the left, a 'Comments' sidebar is visible with a search bar and a dropdown menu. The main area is split into two columns. The left column contains the original Greek text: **Φύλαξ** θεοῦς μὲν αἰῶ τῶνδ' ἀπαλλαγὴν πόνων φρουρὰς ἔτελας μήκος, ἦν κοιόμενος στέγας Ἀτραίδων ἄγκαθεν, κυνὸς δίκην, ἄστρων κάτοιδα νυκτέρων ὀμήγυριν, καὶ τοῦς φέροντας χεῖμα καὶ θέρος βροτοῖς 5 λαμπροῦς δυνάστας, ἐμπρέποντας αἰθέρι ἀστέρας, ὅταν φθίνασιν, ἀντολάς τε τῶν.

The right column contains the Italian translation: **SCOLTA:** Numi, il riscatto concedete a me dei miei travagli, della guardia lunga un anno già, ch'io veglio sui tetti degli Atridi, prostrato su le gonima a mo' d'un cane. E de le stelle veggo 5 il notturno concilio, ed i signori riscaltranti che nell'ètra fulgono, ed il verno e la state all'uomo recano.

Below the text, there are two 'text selection' buttons. A rich text editor is active, showing a selected blue bar above the text 'locuzione avverbiale: come un cane.' The editor includes a toolbar with various formatting options like bold, italic, underline, and text color. At the bottom, there are 'new', 'delete', and 'submit' buttons, along with a 'clear' button.

Fig. 12.2. Annotazioni al testo in lingua originale e alla traduzione

¹⁷ Questa metodologia per la creazione di apparati minimi è ispirata a *Musisque Deoque*, consultabile online all'indirizzo <http://www.mqdq.it> [ultima visita 1.3.2013].

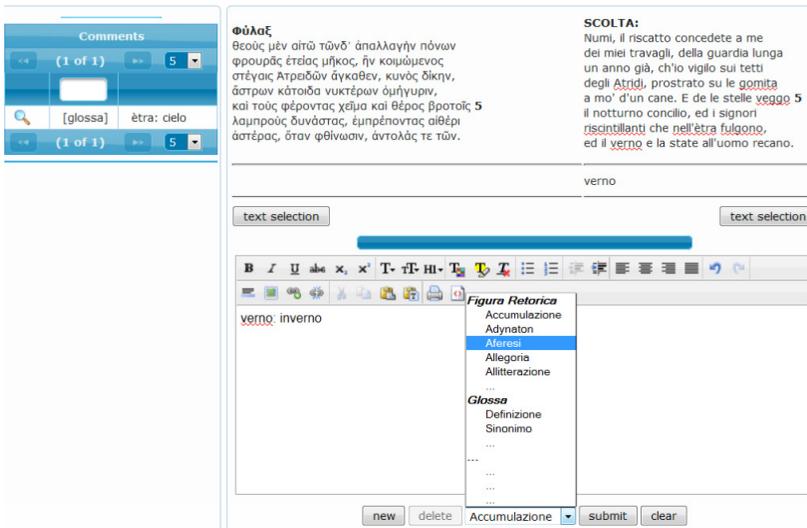


Fig. 12.3. Classificazione delle annotazioni

12.7. L'allineamento a granularità più fine

L'allineamento parola per parola occuperà la seconda fase del progetto e sarà eseguita secondo le metodologie già utilizzate presso il Perseus Project per l'allineamento degli originali alle traduzioni in lingua inglese¹⁸. Data una quantità sufficiente (circa 500.000 parole) di testi allineati a granularità più grezza (almeno pagina per pagina), Giza++¹⁹, software ampiamente usato su corpora plurilingui, ad esempio dei documenti ufficiali del parlamento europeo, è in grado di effettuare l'allineamento parola per parole (quando possibile) o locuzione per locuzione.

Lo scopo dell'allineamento a granularità più fine è duplice. Come primo obiettivo, si intende creare un lessico dinamico bilingue che sia di supporto all'opera di localizzazione del Perseus Project e di ausilio alle attività didattiche summenzionate. Purtroppo, per la lingua italiana non esiste ancora uno strumento simile al Liddell-Scott James

¹⁸ Per i lavori di allineamento e per il loro impiego a scopo di ricerca, si veda Bamman (2008).

¹⁹ Per bibliografia aggiornata su questo strumento e relative proposte di miglioramenti si vedano Pianta *et al.* (2004) e Mehdad *et al.* (2011).

digitale e marcato in T.E.I. consultabile online, a causa dei diritti d'autore sulle risorse cartacee disponibili. I lessici bilingui in formato digitale sono il primo passo per poter compiere studi di letteratura comparata, per individuare, con strumenti informatici e metodi statistici, la persistenza del classico nella letteratura moderna.

Il secondo obiettivo riguarda la puntuale analisi morfosintattica dei testi con traduzione, gestiti tramite l'applicazione sviluppata dall'Alpheios Project, di cui si parla nella prossima sessione.

12.8. Infrastrutture ospiti

Alpheios. L'Alpheios Project²⁰, in stretta collaborazione con il Perseus Project, offre gli strumenti per la visualizzazione e la ricerca di testi in parallelo a granularità di parola, con analisi morfosintattiche che seguono le regole della grammatica di dipendenza. I poemi omerici, le tragedie di Eschilo e di Sofocle, l'*Eutifrone* di Platone, parti dell'*Eneide* ed alcuni testi arabi sono già disponibili; parti dell'*Odissea* sono disponibili anche con l'allineamento parola per parola con la traduzione in lingua inglese.

La piattaforma di Alpheios permette l'accesso a qualunque utente desideri partecipare all'iniziativa di annotazione. Secondo la procedura standard, almeno due annotatori operano sullo stesso testo ed un terzo interviene per l'armonizzazione dei casi discordanti²¹. Secondo la procedura per esperti, invece, una singola opera è affidata ad uno specialista²².

Philologist. Philologist, derivato da Son of Suda OnLine²³ (SoSOL) è l'ambiente per l'editing collaborativo con controllo di versione sviluppato presso il Perseus Project. Le annotazioni insistono su porzioni di testo descritte tramite i Canonical Text Services²⁴ (CTS), che

²⁰ <http://alpheios.net> [ultima visita 1.3.2013].

²¹ Con questo metodo i poemi omerici sono stati annotati da studenti universitari.

²² Questo metodo è stato applicato alle opere di Eschilo e Sofocle da F. Mambrini.

²³ <http://wiki.digitalclassicist.org/SoSOL> [ultima visita 1.3.2013].

²⁴ Per approfondimenti si veda <http://chs75.chs.harvard.edu/projects/diginc/techpub/cts> [ultima visita 1.3.2013].

permettono di identificare in modo univoco, tramite un URI, ciascuna parola di una specifica edizione. In questo modo, tutte le informazioni riferite ad una unità lessicale possono essere agevolmente associate in stand-off.

12.9. Interoperabilità

La collaborazione dell'Istituto di Linguistica Computazionale di Pisa con il Perseus Project ha come fine principale l'interoperabilità delle risorse. La netta separazione fra dati, marcati in T.E.I. e dunque in formato standard, e applicazioni che li gestiscono, promuove la coesistenza di applicazioni diverse che insistono sulle medesime risorse testuali.

Questa politica di collaborazione fra l'Italia e gli Stati Uniti ha già dato risultati visibili, nel recente passato, con la fusione della collezione di testi di Poeti d'Italia in Lingua Latina coordinata da P. Mastandrea nella collezione americana. Grazie al formato standard e agli accordi sulle licenze, la collezione è ora accessibile sia dal portale di Musisque Deoque²⁵ per ricerche intertestuali e metriche, sia dal portale del Perseus Project, per ricerche lemmatizzate.

Sia che i testi siano annotati tramite *Aporia*, sia che vengano annotati tramite *Philologist*, l'interoperabilità è garantita dall'impiego di un robusto sistema di riferimento alle fonti primarie che, in modo diretto o tramite processi di conversione, si basa sui suoi CTS. La stand-off annotation, inoltre, permette di estendere in modo indefinito i livelli di analisi: analisi metrica (sia dell'originale che della traduzione, quando pertinente), analisi colometrica etc. possono essere eseguite in fasi successive e/o da gruppi indipendenti.

L'interoperabilità naturalmente non si ferma soltanto ai dati ma coinvolge anche le applicazioni. Come i dati sono distribuiti con licenza Creative Commons, così il software è distribuito, tanto dal Perseus Project quanto dai suoi collaboratori, fra cui, appunto, il gruppo di Filologia Computazionale dell'ILC, con licenza GPL²⁶.

²⁵ <http://www.mqdq.it> [ultima visita 3.3.2013]

²⁶ <http://www.gnu.org/licenses/gpl.html> [ultima visita 3.3.2013]

12.10. CoPhiLibs

Proprio per garantire interoperabilità e riusabilità, il gruppo di Filologia Computazionale dell'ILC sta sviluppando componenti implementati in Java e organizzati in librerie, al servizio della filologia collaborativa²⁷.

Riusando i medesimi componenti, sono già state realizzate applicazioni in campi diversi, per quanto affini. *G2A* è la web application di supporto al progetto *Greek into Arabic*²⁸ per ricercare e visualizzare testo greco e traduzione araba della *Pseudo-teologia* di Aristotele. *Parallel RGDA*²⁹ mostra testo epigrafico latino e traduzione greca in parallelo, con particolare attenzione alla ricerca e visualizzazione delle parte integrate congetturalmente. *Saussure WA* sostituisce invece ad un componente di visualizzazione testuale un componente di visualizzazione di immagini, mostrando, in parallelo, l'immagine della pagina scritta da Saussure e la relativa trascrizione. È inoltre in corso di sviluppo uno strumento di ausilio alla correzione dell'OCR e (con tecnologia indipendente dalle CoPhiLibs) un sistema di traduzione collaborativa, il Progetto Talmud.

12.11. Conclusioni

In conclusione, il progetto di localizzazione del Perseus Project in lingua italiana ha stimolato il gruppo di Filologia Computazionale dell'Istituto di Linguistica Computazionale del CNR di Pisa a proseguire la via intrapresa per la creazione di risorse aperte ed interoperabili, che possano essere aggregate in molteplici progetti italiani e internazionali e che possano essere ospitate in infrastrutture solide e scalabili come quelle offerte dal Perseus e dai suoi collaboratori.

²⁷ Le CoPhiLibs, per quanto ancora ad uno stadio aurorale, sono disponibili all'indirizzo <http://github.com/CoPhi>.

²⁸ Si veda Bozzi (2013).

²⁹ Si veda Boschetti (2013).

12.12. Bibliografia

- ACRI F., *trad. a c. di* (1913-15). *Platone, Dialoghi*. Vallecchi.
- BABEU A. (2011). *Rome Wasn't Digitized in a Day: Building a Cyber-infrastructure for Digital Classicists*. Council on Library and Information Resources. URL= <http://www.clir.org/pubs/reports/pub150> [ultima visita 19.2.2013].
- BAMMAN D. - CRANE G. (2008). *The Logic and Discovery of Textual Allusion*. In *LaTeCH (Language Technology for Culturale Heritage Data) 2008 Marrakech Morocco*. URL= <http://hdl.handle.net/10427/42685> [ultima visita 19.2.2013].
- BATTISTA M. - PIRRELLI V. (1999). *Una piattaforma di morfologia computazionale per l'analisi e la generazione delle parole italiane*. ILC-CNR technical report.
- BOSCHETTI F. (2013). *Strumenti per l'analisi di testi bilingui al servizio dell'epigrafia digitale*. Articolo sottoposto a *peer review* presso la redazione di «Lexis».
- BOZZI A. (2013). *G2A: a web scholarly application for texts, sources and translations*. «*Studia graeco-arabica*» vol. 3 no 1, in corso di pubblicazione.
- CRANE G. *et al.* (2012). *Student researchers, citizen scholars and the trillion word library*. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (JCDL '12)*. ACM. URL=<http://doi.acm.org/10.1145/2232817.2232857> [ultima visita 19.2.2013].
- DELL'ORLETTA F. *et al.* (2007). *La massima entropia per il Part of Speech Tagging dell'Italiano - Maximum Entropy for Italian POS Tagging*. In *Proceedings of Workshop Evalita 2007*. «Intelligenza artificiale», vol. 4, no 2.
- DELL'ORLETTA F. (2009). *Ensemble system for Part-of-Speech tagging*. In *Proceedings of Evalita '09*. URL= http://www.evalita.it/sites/evalita.fbk.eu/files/proceedings2009/PoSTagging/POS_ILC.pdf [ultima visita 19.2.2013].
- MEHDAD Y. *et al.* (2011). *Using bilingual parallel corpora for cross-lingual textual entailment*. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11)*, Vol. 1. Association for Computational Linguistics, pp. 1336-1345. URL= <http://www.aclweb.org/anthology-new/P/P11/P11-1134.pdf> [ultima visita 19.2.2013].
- NORTH FOWLER H., *trad. a c. di* (1913). *Plato with an English translation*. Loeb.

- PIANTA E. - BENTIVOGLI L. (2004). *Knowledge intensive word alignment with KNOWA*. In *Proceedings of the 20th international conference on Computational Linguistics (COLING '04)*. Association for Computational Linguistics, article 1086. URL= <http://dx.doi.org/10.3115/1220355.1220511> [ultima visita 19.2.2013].
- ROMAGNOLI E., *trad. a c. di* (1922). *Eschilo, Le tragedie: Agamennone, Le Coefore, Le Eumenidi*. I poeti greci tradotti da E. Romagnoli, vol. II. Zanichelli.
- SMYTH H.W., *trad. a c. di* (1926). *Aeschylus with an English translation by H.W. Smyth, Ph.D. in two volumes, vol. II: Agamemnon*. Harvard University Press.

13. Il progetto ENArC. Attività didattiche innovative e creazione di archivi digitali

Maria Rosaria Falcone¹

Abstract. L'articolo presenta le attività del Dipartimento di Studi Umanistici dell'Università degli Studi di Napoli Federico II nell'ambito del progetto internazionale ENArC (European Network on Archival Cooperation - EU, Culture Programme 2007-2013) che ha come finalità primaria la creazione e la diffusione di una rete europea di istituzioni per la tutela, la conservazione e la fruibilità della loro documentazione, quale patrimonio comune della cultura scritta.

Il gruppo di lavoro dell'Università di Napoli è impegnato in due ambiti di ricerca e di azione:

- la realizzazione di una didattica innovativa per la Paleografia e la Diplomatica con l'impiego delle nuove tecnologie.
- la creazione di archivi digitali, contenenti documentazione conservata in Italia e in Spagna, per il portale Monasterium.Net, nel quale sono presenti, oggi, più di 230.000 documenti e quasi 500.000 riproduzioni, descritti grazie ad un software per la gestione dei documenti medievali, MOM-Ca, implementato dall'Università di Colonia specificamente per il progetto.

Parole chiave: didattica, nuove tecnologie, cooperazione internazionale, collaborazioni interdisciplinari, monasterium.net, documenti storici.

¹ Dipartimento di Studi Umanistici, Università degli studi di Napoli Federico II, Napoli, Italia. e-mail: mariarosaria.falcone@unina.it.

13.1. Il progetto ENArC- European Network on Archival Cooperation. Un esempio di collaborazioni interdisciplinari e internazionali

Presso il Dipartimento di Studi Umanistici dell'Università degli Studi di Napoli Federico II, nell'ambito delle attività progettuali finanziate dalla Comunità Europea per il Programma Cultura 2007-2013, è attivo dal 2010 (con scadenza il 15 aprile 2015), il progetto internazionale ENArC -European Network on Archival Cooperation² le cui attività rappresentano un esempio di collaborazione interdisciplinare a carattere internazionale e non solo accademico.

Il progetto nasce dalla sinergia e dalla cooperazione di enti internazionali e coinvolge diciannove istituzioni, in particolare, archivi, università, istituzioni scientifiche che operano in undici Paesi europei.³

L'obiettivo fondamentale è di creare ed ampliare una rete istituzionale europea che tuteli e promuova la conservazione e la fruizione della documentazione, quale patrimonio comune della cultura scritta.

Il gruppo di lavoro dell'Università di Napoli, coordinato da Antonella Ambrosio, docente di Paleografia⁴, si impegna, in tale contesto, in più direzioni.⁵

² L'indirizzo del progetto: <http://enarc.icar-us.eu/>. [ultima visita 28.02.2013].

³ I partner del progetto sono: Diözesanarchiv St.Pölten (AT); ICARUS – International Centre for Archival Research (AT); Magyar Nemzeti Levéltár (HU); Budapest Főváros Levéltárában (HU); Ministerstvo vnútra Slovenskej republiky (SK); Národní Archiv (CZ); Arhiv Republike Slovenije (SI); Hrvatski državni arhiv (HR); БАЛКАНОЛОШКИ ИНСТИТУТ СРПСКА АКАДЕМИЈА НАУКА И УМЕТНОСТИ (RS); Die Staatlichen Archive in Bayern (DE); Università degli Studi di Napoli Federico II (IT); Scuola Normale Superiore di Pisa (IT); Sub-Directorado General de los Archivos Statales (ES); Universität zu Köln (DE). I Partner associati: Institute for Indexing and Research of Clerical Sources (AT); Archiv des Bistums Passau (DE); Moravský zemský archiv v Brně (CZ); Österreichische Staatsarchiv (AT); Stadtarchiv Speyer (DE). È possibile accedere ai rispettivi siti dall'indirizzo: <http://www.recruitdigitaldoc.org/> [ultima visita 28.02.2013].

⁴ Si veda per i riferimenti: <http://www.docenti.unina.it/antonella.ambrosio> [ultima visita 28.02.2013].

⁵ Per le quali è possibile consultare il sito del progetto, alla pagina: http://www.recruitdigitaldoc.org/?page_id=118 [ultima visita 28.02.2013].

L'aspetto caratterizzante delle attività di questo gruppo di lavoro è costituito dalle collaborazioni interdisciplinari ed in qualche modo trasversali tra le istituzioni con cui è in contatto. Molte di queste attività convergono col progetto *Monasterium.Net*⁶ (Heinz 2009, 40-55; Krah 2009, 221-246; Aigner 2010, 123-128; Ambrosio 2010, 129-145).

Monasterium, attualmente, si presenta come una delle risorse documentarie europee più ingenti disponibili nel Web, sia per quanto riguarda i documenti medievali sia per quelli risalenti all'Età moderna e, di recente, anche per la documentazione d'Età contemporanea. Il portale contiene un numero di riproduzioni digitali di documenti, circa cinquecentomila unità, riprodotti ad altissima risoluzione. Tale elevato numero di riproduzioni è contenuto in centoquattro archivi digitali, prodotti da tredici Paesi europei⁷.

Le fotografie ed i dati ad esse correlati, sono accessibili sul portale *Monasterium.Net*, attraverso il software MOM-Ca implementato dall'Università di Colonia appositamente per la gestione dei documenti medievali e basato sugli standard di descrizione internazionali (CEI). Il software dispone di un editor XML, basato sul linguaggio di programmazione Java, chiamato Edit MOM (Burkard et al. 2007, 193-210, Burkard 2007, 130-144; Burkard 2009, 255-27).

Il relativo ambiente di *editing* risulta di utilizzo semplice ed intuitivo: è possibile, infatti, descrivere in maniera agevole e, allo stesso tempo, molto dettagliata i documenti, grazie alle *tab* dedicate all'inserimento del regesto, della trascrizione, della descrizione dei caratteri estrinseci del testimone (originale, copie ecc.).

⁶ L'indirizzo del progetto: <http://www.monasterium.net> [ultima visita 28.02.2013].

⁷ I paesi rappresentati in *Monasterium.Net*, con dati aggiornati al febbraio 2013, sono: Austria, Croazia, Germania, Italia, Polonia, Repubblica Ceca, Romania, Serbia, Slovacchia, Slovenia, Spagna, Svizzera, Ungheria. È possibile consultare gli archivi digitali pubblicati al seguente indirizzo: <http://www.mom-ca.uni-koeln.de/mom/fonds> [ultima visita 28.02.2013].

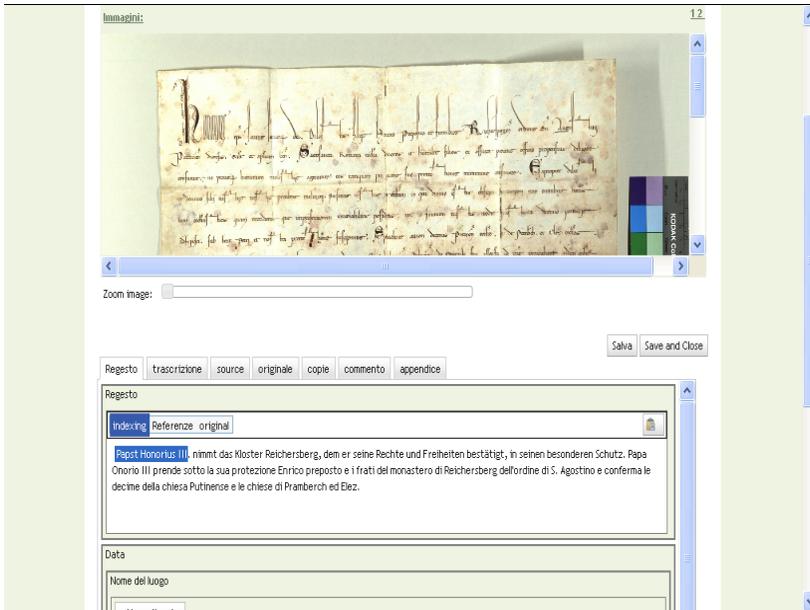


Fig. 13.1. Edit MOM Tool. Inserimento del regesto.

Edit MOM inoltre consente, una volta terminato l'inserimento dei dati all'interno di ciascuna *tab*, di effettuare su di essi la cosiddetta operazione di *marcatura*, che permette di espletare in modo agevole una vera e propria operazione di codifica XML. Si tratta di fornire delle informazioni aggiuntive che specificano ulteriormente il dato inserito e che vengono catturate dal motore di ricerca presente in Monasterium, semplificando, in tal modo, la consultazione della documentazione degli archivi digitali.

EditMOM, grazie al Web 2.0, rende possibile un'**attività collaborativa** anche, e soprattutto, a distanza.

Una tipologia, questa ultima, di attività simile, nel principio, a quella attuabile in Wikipedia. Naturalmente i dati, per essere resi accessibili online, devono essere approvati e pubblicati da un **moderatore**, un esperto di Archivistica e di Diplomatica che si assume la responsabilità di garantire la qualità scientifica dei dati immessi dagli utenti, che possono anche essere studenti.

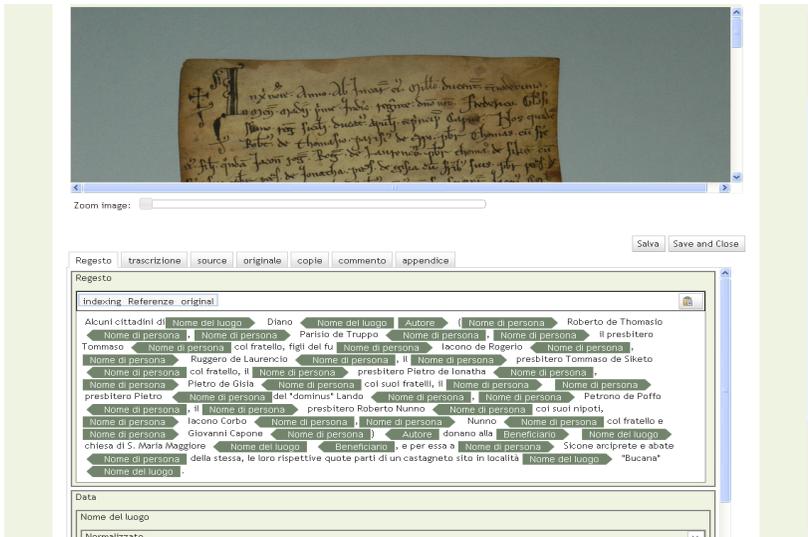


Fig. 13.2. Edit MOM Tool. La marcatura XML.

L'attività realizzata in MOM-Ca risponde, nella sostanza, a tre principi fondamentali: anzitutto, al recupero di un'enorme mole di informazioni immediatamente fruibili nelle sedi di conservazione, ma, molto spesso, difficilmente consultabili a distanza, nell'ottica di un continuo e progressivo arricchimento. In secondo luogo, si propone di garantire la libera circolazione del sapere a tutti gli interessati, i quali possono accedere ed effettuare liberamente il *download* delle immagini ad alta risoluzione e consultare i dati correlati ai documenti presenti online. In ultimo, il rispetto, sebbene in un obiettivo e in una prospettiva internazionale, delle differenti modalità di descrizione archivistica e di edizione dei documenti e dell'utilizzo delle relative lingue europee.

Intorno all'utilizzo del portale Monasterium.Net e del software MOM-Ca, ruotano molte attività del gruppo di lavoro dell'Università di Napoli. Prima su tutte, le **attività didattiche**.

Presso l'Ateneo, infatti, sin dal 2008, Antonella Ambrosio, tiene i corsi di Paleografia e Diplomatica per il corso di Laurea Magistrale in Scienze Storiche e di Diplomatica per il Master in Biblioteconomia e Archivistica con l'ausilio di MOM-Ca Software e di Edit MOM Tool.

I corsi sono strutturati in tre fasi: una fase preliminare, costituita da un'introduzione alle discipline ed alla metodologia tradizionale di edizione di documenti. Una intermedia, durante la quale la docente

guida gli studenti nel passaggio dalla metodologia tradizionale a quella digitale introducendoli, in primo luogo, alle risorse online esistenti e consultabili. Infine, la terza fase, è dedicata alle attività laboratoriali svolte in aula informatica, precedute da un'introduzione al linguaggio di codifica XML ed alla pratica di edizione con Edit MOM Tool.

All'interno del dipartimento di Studi Umanistici, relativamente alle attività didattiche ma anche di ricerca, si è costituita nel 2011 un'équipe interdisciplinare formata da Antonella Ambrosio (docente di Paleografia e Diplomatica), Maura Striano (docente di Pedagogia e responsabile del progetto FIRB Rete@ccessibile⁸), Corinna Freda (ingegnere biomedico), Stefania Fiorentino (pedagogista). La collaborazione tra i membri dell'équipe ha permesso la sperimentazione di corsi di Diplomatica, presso il Master e presso il Corso di Laurea in Scienze Storiche, con l'ausilio delle tecnologie digitali fin qui descritte combinate con la piattaforma Moodle utilizzata nel progetto Rete@ccessibile.

Dall'approccio interdisciplinare sono scaturite riflessioni che hanno condotto, nel 2011, alla partecipazione alla conferenza internazionale *Digital Diplomatics. Tools for the Digital Diplomatist*, tenutasi presso l'Università degli studi di Napoli Federico II e presso la Società Napoletana di Storia Patria, durante la quale Antonella Ambrosio e Maura Striano hanno presentato una relazione dal titolo *Teaching Diplomatics with New Technologies: a Possible Path*. A questo primo convegno è seguita la partecipazione alla undicesima conferenza europea sull'e-Learning, tenutasi presso l'Università di Groningen, nell'ottobre del 2012 nell'ambito della quale l'intero gruppo ha presentato una relazione dal titolo *Teaching in a Web 2.0 environment Diplomatics: an innovative experience of internationalization*. Il focus di questo intervento è stato centrato sulla presentazione dell'esperienza di insegnamento con le nuove tecnologie (MOM-Ca, Rete@ccessibile) e su una riflessione sul potenziamento delle possibilità di apprendimento che ne deriva. In accordo con una prospettiva teorica "costruttivista" è stato, infatti, allestito un ambiente di apprendimento adatto alle potenzialità dei singoli studenti (Ambrosio et al. 2012, 16-17).

⁸ Per la descrizione ed i dettagli relativi al progetto FIRB Rete@ccessibile, si veda: <http://www.firbreteaccessibile.it/Progetto.html> [ultima visita 28.02.2013].

Le attività attualmente in essere riguardano, inoltre, le collaborazioni internazionali. Nell'ambito del progetto, e col sostegno di uno dei partner, ICARus-International Centre for Archival Research⁹, si è costituito l'ICARus Didactics Group, coordinato dall'Università di Napoli proprio per l'esperienza maturata in questi anni nell'ambito delle attività didattiche innovative. Il gruppo si è costituito nel giugno del 2011, in occasione dell'ICARus Meeting tenutosi a Poznań (PL), ed è attualmente composto da ventotto membri (insegnanti e ricercatori di Archivistica, Paleografia, Diplomatica, Scienze ausiliari e Storia Medioevale) provenienti da ben quattordici paesi europei, e da tre tutor¹⁰. Fra i suoi componenti un più ristretto numero di docenti del settore lavora sull'impianto teorico e sulla metodologia di insegnamento con il supporto delle nuove tecnologie. Il gruppo si riunisce semestralmente in un workshop, in occasione dell'ICARus Meeting¹¹. Uno degli obiettivi su cui si sta attualmente impegnando è la progettazione di un corso internazionale di Diplomatica a distanza, che sfrutti il software MOM-Ca.

In particolare, il gruppo di docenti sta elaborando le unità didattiche che potrebbero far parte del corso internazionale, che sarà poi fruibile su una piattaforma Moodle. Le prime proposte sono state presentate durante il Workshop svoltosi ad Einsiedeln (CH), nel giugno del 2012.¹²

13.2. Attività interdisciplinari trasversali: lo *StartUp*

⁹ L'indirizzo di ICARus: <http://icar-us.eu/> [ultima visita 02.03.2013].

¹⁰ I paesi finora coinvolti sono: Austria, Bulgaria, Croazia, Finlandia, Francia, Germania, Italia, Malta, Polonia, Repubblica Ceca, Serbia, Slovacchia, Spagna, Ungheria.

¹¹ L'elenco dei workshop tenuti, a partire dal 2011, è consultabile alla pagina: http://www.recruitdigitaldoc.org/?page_id=391 [ultima visita 02.03.2013].

¹² Durante il workshop sono intervenuti: Nicolás Avila Seoane (Universidad Complutense de Madrid), Susana Cabezas Fontanilla (Universidad Complutense de Madrid), Adele Di Lorenzo (Università di Siena/Arezzo), Francesca Klein (Archivio di Stato di Firenze), Dominique Stutzmann (Institut de Recherche et d'Histoire des Testes -Paris), il quale ha presentato anche le proposte dei seguenti studiosi: Olivier Guyotjeannin (École Nationale des Chartes), Paul Bertrand (Institut de Recherche et d'Histoire des Textes) e Sébastien Barret (Institut de Recherche et d'Histoire des Textes).

degli archivi digitali in Monasterium.Net

Nell'ambito delle attività progettuali, il gruppo di lavoro dell'Università di Napoli promuove la creazione di archivi digitali per Monasterium.Net. A tale scopo, grazie alla collaborazione con il gruppo di Monasterium che si occupa dello sviluppo del software¹³, e nell'ambito delle attività didattiche, l'Università di Napoli ha avviato un'attività di affiancamento agli enti che intendono pubblicare un archivio digitale in Monasterium.Net.

Tale attività si realizza nel trasferimento agli enti di competenze specifiche (linguaggio di codifica XML, creazione dei metadati, digitalizzazione) attraverso un percorso di formazione a distanza e in presenza, che prevede anche attività di *Learning by doing* volte alla sperimentazione del software MOMCa e del tool Edit MOM, su esempi concreti tratti dagli archivi digitali degli enti stessi.

L'intervento del gruppo di lavoro di Napoli si conclude, salvo necessità particolari, quando gli archivi vengono pubblicati online.

In tal senso, l'Università Federico II sta curando la fase di *startup* di alcuni archivi digitali italiani: l'Archivio Storico Diocesano di Reggio Calabria-Bova, che recentemente ha pubblicato sul portale il proprio fondo diplomatico¹⁴ (Mazzitelli 2013), e l'Archivio di Stato di Potenza che ha, invece, avviato il lavoro di digitalizzazione del proprio patrimonio pergameneo. Il gruppo di lavoro di Napoli si sta, inoltre, occupando di alcuni archivi digitali internazionali: l'archivio Storico Municipale di Escalona¹⁵, e l'Archivio del Capitolo della Cattedrale di Cuenca¹⁶. Riguardo alla creazione di questi due archivi spagnoli, è utile sottolineare che è attualmente in corso una collaborazione tra l'Italia e

¹³ Il Monasterium Technical Development, costituito, all'interno di ICARus, come gruppo di lavoro, è composto da Georg Vogeler, (University of Graz -AT), che ne è il coordinatore e da Manfred Thaller (University of Cologne -DE); Jochen Graf (University of Cologne -DE); Karl Heinz (ICARus); Daniel Jeller (ICARus). I riferimenti all'indirizzo: <http://icar-us.eu/activities/didactic-work-groups/members/> [ultima visita 02.03.2013].

¹⁴ Al seguente indirizzo è consultabile l'archivio digitale: <http://www.mom-ca.uni-koeln.de/mom/IT-ASDRCB/archive> [ultima visita 02.03.2013].

¹⁵ I futuri curatori: Susana Cabezas Fontanilla e Nicolás Avila Seoane -Universidad Complutense de Madrid.

¹⁶ I futuri curatori: Francisco Antonio Chacón Gómez-Monedero -Universidad Autonoma de Madrid; Manuel Joaquín Salamanca López-Universidad Complutense de Madrid.

la Spagna nell'ambito del programma speciale afferente al progetto ENArC, denominato *Linking The Neighborhood*¹⁷, un programma che prevede la collaborazione e lo scambio con paesi vicini geograficamente e/o culturalmente affini, come nel caso, appunto, dell'Italia e della Spagna.

13.3. La didattica e gli archivi digitali come spunti di ricerca

Gli ambiti di azione del gruppo di lavoro di Napoli, vale a dire, la sperimentazione didattica con le nuove tecnologie e la promozione della creazione degli archivi digitali in Monasterium.Net, sono entrambi direttamente connessi ad una attività di ricerca che attualmente è in corso presso il dipartimento di Studi Umanistici.

Per quanto riguarda la ricerca in ambito didattico, essa è, naturalmente, collegata alle attività del Didactics Work Group e alle collaborazioni interdisciplinari, già citate, che stanno incentivando una riflessione profonda sulle possibilità di apprendimento potenziato dalla tecnologia.

Per quanto concerne gli archivi digitali, l'Università di Napoli Federico II promuove direttamente la creazione di archivi digitali che, all'interno del progetto ENArC, si configurano come veri e propri *sub-project*. Questa attività ha capacità trasversali di coinvolgere ricercatori, studenti e tirocinanti. Il gruppo di lavoro, infatti, al fine di promuovere la formazione di un esperto della documentazione, le cui competenze spazino da una metodologia tradizionale a quella digitale, ha, nel corso del 2011, organizzato una serie di workshop dedicati al concetto di archivio digitale come *unicum* composto da riproduzione digitale del documento e dati connessi, alla Codifica XML ed all'utilizzo del software MOM-Ca. Alcuni partecipanti hanno fattivamente collaborato alla creazione di archivi digitali, tra i quali, l'archivio digitale della

¹⁷ La descrizione del programma e i paesi che, nell'ambito del progetto ENArC, sono coinvolti, alla seguente pagina: http://www.recruitdigitaldoc.org/?page_id=448 [ultima visita 02.03.2013].

*Biblioteca della Società Napoletana di Storia Patria*¹⁸.

Per la realizzazione di questo archivio, l'Università di Napoli, con il sostegno economico di ICARus, ha promosso ed effettuato la digitalizzazione dell'intero posseduto pergameneo, di circa tremila documenti, della biblioteca e la creazione dei dati minimi con il coinvolgimento di giovani tirocinanti appassionati a questa attività¹⁹.

Un ulteriore progetto di ricerca applicata sugli archivi digitali è quella che, attualmente, sta conducendo Rosalba Di Meglio, ricercatrice e docente di storia medievale, per l'archivio digitale del *Monastero di S. Agostino Maggiore*²⁰ di Napoli, già pubblicato in *Monasterium.net* con fotografie e dati minimi. In particolare, l'archivio accoglierà i risultati della ricerca condotta dalla studiosa per i secoli XIV-XV, includendo l'edizione critica dei relativi documenti. Ad una fase intermedia del progetto ha partecipato anche una tirocinante²¹, la quale, sulla base dei dati elaborati dalla ricercatrice, ha effettuato la codifica degli stessi per la futura pubblicazione.

Per la sua natura collaborativa, è, inoltre, particolarmente utile sottolineare la nascita di un sub-project dal titolo *The Digital Archive of Late Middle Ages Women's Convents in Naples*, con l'obiettivo di creare gli archivi digitali in *Monasterium.Net* di tutti i monasteri femminili medievali della città di Napoli. I documenti sopravvissuti provenienti dagli archivi di questi monasteri coprono i secoli IX-XV. Le pergamene o anche i documenti cartacei relativi saranno pubblicati sul portale *Monasterium.Net*, mentre l'edizione critica dei registri amministrativi sarà pubblicata online, ma secondo forme e modalità ancora in fase di definizione. Per lo sviluppo di questo sub-project si è costituito un gruppo di lavoro coordinato da Antonella Ambrosio e formato da Valentina Di Girolamo (PhD in Filologia), Francesco Li Pira (PhD in Storia), Ciro Romano (PhD student in

¹⁸ Attualmente, di questo archivio è consultabile il fondo *Carte Feudali*: <http://www.mom-ca.uni-koeln.de/mom/IT-BSNSP/CarteFeudali/fond> [ultima visita 03.03.2013].

¹⁹ La digitalizzazione e la codifica dei dati in XML sono stati coordinati da Antonella Ambrosio. Per la digitalizzazione, Michele Innangi. Per i dati: Maria Rosaria Falcone, tutor; Olimpia Di Palo, Angelica Parisi, Concetta Prisco, Pasqualina Uccello (laureate in Lettere classiche e moderne), tirocinanti.

²⁰ L'archivio, in *progress*, è consultabile: <http://www.mom-ca.uni-koeln.de/mom/IT-ASNA/ASAg/fond> [ultima visita 03.03.2013].

²¹ Viviana De Luca, laureata in Storia e con Master in Archivistica e Biblioteconomia.

Cultura Italiana) e Maria Rosaria Falcone (PhD student in Storia).

Sembra, in conclusione, di poter affermare che l'utilizzo delle tecnologie del web 2.0, su cui si basa la correlazione tra tutte le attività finora descritte, consenta, attraverso la sperimentazione continua, di costruire, grazie alla loro naturale predisposizione all'*interoperabilità*, una dimensione del "comune" in cui si esprimono competenze trasversali e complementari, che, in quanto tali, permettono di dare luogo ad attività ed esperienze collaborative multidisciplinari capaci di aprire prospettive di apprendimento e di ricerca che favoriscono la formazione di una coscienza comune del patrimonio culturale europeo.

13.4. Bibliografia

- AIGNER T. (2010). *Monasterium.Net-Documenti Europei online*. «Archivi» vol. V, no 2, pp. 123-128.
- AMBROSIO A. (2010). *Il progetto 'monasterium' in Italia: le prime esperienze a Napoli*. «Archivi» vol. V, no 2, pp. 129-145.
- AMBROSIO A. (2011). *Towards the Creation of a Learning Environment within the Monasterium Project: Teaching Experiences of Diplomats*. In Aigner T., Hohenbruck S., Just T., Kemper J., a c. di, *Proceedings of Archives on the Web - Experiences, Challenges, Visions*, Diözesanarchiv St. Pölten, pp. 203-215.
- AMBROSIO A. (2012). *Insegnare la diplomazia con le nuove tecnologie. Potenzialità e spunti di riflessione*. In Cherubini P., Nicolaj G., a c. di, *Sit liber gratus, quem servulus est operatus*. Studi in onore di Alessandro Pratesi per il suo 90° compleanno, Città del Vaticano, Littera Antiqua, 19, II, pp. 1315-1326.
- AMBROSIO A. et al. (2012). *Teaching Diplomats in 2.0 Web Environments: An Innovative Experience to Promote Interaction Among Students From Different Countries and With Different Learning Needs*. In Beldhuis H., a c. di, *The Proceedings of the 11th European Conference on Learning*, University of Groningen, The Netherlands, pp. 15-19.

- BURKARD B., VOGELER G., *et al.* (2008), *Informatics for Historians: Tools for Medieval Document XML Markup, and their Impact on the History-Sciences*. «Journal of Universal Computer Science», vol.14, no 2, pp. 193-210. URL= http://www.jucs.org/jucs_14_2/informatics_for_historians_tools/jucs_14_2_0193_0210_burkard.pdf [ultima visita 01.03.2013].
- BURKARD B. (2007), *Wiki goes Humanities. Kollaborative Erschließung mittelalterlicher Urkunde*. In Stockinger J., Leitner H., a c. di, «Wikis im Social Web – Wikiposium», 2005/06, pp. 130-144.
- BURKARD B. (2009). *EditMOM - ein spezialisiertes Werkzeug zur kollaborativen Urkunden-Erschließung*. In Vogeler G., a c. di, *Digitale Diplomatie. Neue Technologien in der historischen Arbeit mit Urkunden*. In «Archiv für Diplomatie», no 12, pp. 255-27.
- HEINZ K. (2009). *Monasterium.net: Auf dem Weg zu einem mitteleuropäischen Urkundenportal*. In Vogeler G., a c. di, *Digitale Diplomatie. Neue Technologien in der historischen Arbeit mit Urkunden*. In «Archiv für Diplomatie», no 12, pp. 40- 55.
- KRAH, A. (2009), *Möglichkeiten der Bereitstellung und Erschließung von Urkundenbeständen*. In «Archivalische Zeitschrift», no 91, pp. 221-246.
- MAZZITELLI M.P. (2013). *Un'esperienza di digitalizzazione all'Archivio storico diocesano di Reggio Calabria-Bova*. «Il Mondo degli Archivi». URL= <http://www.ilmondodegliarchivi.org/index.php/attivita/item/115-un%E2%80%99esperienza-di-digitalizzazione-all%E2%80%99archivio-storico-diocesano-di-reggio-calabria-bova> [ultima visita 02.03.2013].

14. Letteratura latina e ebook. Una prospettiva della biblioteca digitale digilibLT

*Maurizio Lana*¹

Abstract. La biblioteca digitale digilibLT distribuisce liberamente sotto licenza Creative Commons le opere del latino tardo in vari formati tra cui ePUB, il formato aperto degli ebook. Ciò prefigura che lo studioso possa costruire senza difficoltà una biblioteca personale che raccoglie l'intera letteratura. Una volta che ciò avvenga, l'utilizzo di programmi di gestione quali Calibre o iBook permetterà di passare dalla semplice ricerca di passi ad un approccio di studio dall'orizzonte più ampio.

Abstract. The digilibLT digital library distributes freely under Creative Commons license late latin works in various formats, including ePUB which the open format for e-book. It allows to envisage a scholar who builds easily a persona library containing the whole latin literature. Once done, this will allow through the use of computer programs like Calibre or iBook to leave the usual simple search for passages and moving towards a radically different approach.

Parole chiave: accesso aperto, free software, ePUB, Oxygen, PHI, TLG, ebook, licenze Creative Commons, testi latini tardoantichi.

¹ Dipartimento di Studi Umanistici, Università degli Studi del Piemonte Orientale, Vercelli, Italia. e-mail: m.lana@lett.unipmn.it.

14.1. Introduzione

Il progetto digilibLT - digital library of late latin texts / biblioteca digitale di testi latini tardi (<http://www.digiliblt.unipmn.it>) nasce da un'idea di Raffaella Tabacco realizzata con Maurizio Lana, vale a dire creare una biblioteca digitale del latino tardo pagano, in un ideale completamento e continuazione della raccolta di testi latini nota tra gli studiosi di cose classiche come PHI cdrom 5.3. Tale raccolta contiene tutti i testi latini dalle origini fino al I/II sec. d.C., codificati secondo uno standard di fatto noto come Beta code, io medesimo utilizzato per la grande raccolta dei testi greci dalle origini fino all'epoca bizantina, nota come TLG cdrom E. L'idea di completare la raccolta del PHI cdrom con la creazione di una biblioteca digitale del latino tardo venne tradotta in forma di progetto da chi scrive, presentata alla Regione Piemonte nel quadro del "Bando Scienze Umane e Sociali" del 2008 e nel 2009 approvata e finanziata. Il progetto è attualmente in corso presso il Dipartimento di Studi Umanistici dell'Università degli Studi del Piemonte Orientale, e ha come partner il Dipartimento di Studi Umanistici dell'Università degli Studi di Torino².

14.2. Caratteristiche rilevanti della biblioteca digitale

La biblioteca presenta molteplici caratteristiche rilevanti che non possono essere descritte tutte in dettaglio. Ne vengono qui esposte alcune che risultano particolarmente rilevanti nel contesto di una riflessione sulle digital humanities, sull'interazione tra digitale e testuale. Alla base di tutto, la scelta di fornire testi di qualità attraverso un percorso di lavoro di estrema cura: ciò significa che i testi provengono da edizioni critiche reputate valide dal comitato scientifico del progetto (Raffaella Tabacco, Andrea Balbo, Ermanno Malaspina, Roberta Piastri), ma anche che il testo digitalizzato ottenuto da scansione e riconoscimento ottico dei caratteri viene

² Il gruppo di ricerca è costituito da Raffaella Tabacco, Maurizio Lana, Luigi Battezzato, Silvia Botto, Roberta Piastri, Gabriella Vanotti, dell'Università del Piemonte Orientale, e da Andrea Balbo e Ermanno Malaspina dell'Università di Torino; con la collaborazione di David Paniagua dell'Università di Salamanca, di Peter Heslin dell'Università di Durham e di Fabio Ciotti dell'Università di Roma Tor Vergata.

sottoposto a una doppia revisione, la cui validità è confermata dal fatto che sono pochissime le edizioni critiche a stampa in cui non vengano individuati errori di stampa o errori materiali, a riprova del fatto che la biblioteca digilibLT non fornisce semplicemente 'dei file da scaricare', ma vere e proprie *edizioni dei testi* (delle quali è una componente fondamentale l'emendamento degli errori); ma su questo si veda anche sotto.

Panegyrici Latini

Auctores uarii Scheda Autore

saec. III-IV

LEGGI IL TESTO

Download possibili

TXT
TEI
BETA
PDF
E-PUB

SCHEDA CATALOGRAFICA

Edizione di riferimento:

XII Panegyrici Latini, recognouit D. Lassandro, Augustae Taurinorum 1992 (Corpus Scriptorum Latinorum Paruaianum).

Nel 1433, nella biblioteca della cattedrale di Magonza, venne scoperto da Giovanni Aurispa un codice contenente l'orazione di Plinio a Traiano seguita da altri undici discorsi ufficiali presentati tutti nella *inscriptio* come *panegyrici*, tranne la *gratiarum actio* di Claudio Mamertino a Giuliano. Il termine *panegyricus* tuttavia non compare nei testi. Nella tarda latinità (cf. Sidonio Apollinare, *epist.* VIII 10, 3: *Gaius Plinius [...]* *Marco Vlpio incomparabili principii comparabile panegyricum dixit*) il termine *panegyricus* perde il valore originario di discorso tenuto di fronte a un'assemblea universale, per assumere quello di *laudatio*, discorso elogiativo che si

Operazioni preliminari: Nadia Rosso

Codifica TEI: Nadia Rosso

Edizione digitale a cura del gruppo digilibLT - Università degli Studi del Piemonte Orientale

← Torna all'elenco

Bibliografia

1. Camastra P. **Letteratura Latina Tardoantica**
2. Cavallo G., Fedeli P., Giardina A. **Lo Spazio Letterario di Roma Antica. Volume II: la circolazione del testo**
3. Giardina A. - Silvestrini M. **Il principe e il testo**
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**
5. Lassandro D. - Micunco G. **Panegyrici Latini**
6. Lassandro D. I **«Panegyrici Latini» del III-IV secolo**
7. Lassandro D. **XII Panegyrici Latini**
8. Mynors R.A.B. **XII Panegyrici Latini**

1 2 »

Fig. 14.1. Come si presenta un'opera nella biblioteca.

14.2.1. Marcatura TEI/XML

Le opere della biblioteca sono tutte marcate secondo lo standard definito dalla TEI (Text Encoding Initiative) con l'intento di evitare la frammentazione dei modi di marcare nei testi informazioni importanti (per esempi i fenomeni filologici; o la struttura dei testi con la divisione in libri, capitoli, paragrafi; ecc.; tutti aspetti sospesi tra esistenza come concetti e espressione per mezzo di una concreta realizzazione (tipo grafica) e promuovere così la circolazione dei testi. Con un approccio che si potrebbe definire "super light" (alludendo al subset dello

standard noto come TEI light) digilibLT ha fatto una scelta di marcatori estremamente parca in quanto si intende descrivere e/o gestire soltanto:

- struttura dell'opera
- glosse e scoli
- fenomeni filologici (integrazioni, espunzioni, lacune, ...)
- immagini incorporate nelle opere (nel caso di opere di geometria o di musica).

Per ogni opera il TEI header contiene informazioni sull'opera stessa, e in particolare³, come si vede dall'esempio qui sotto:

- titolo, ed autore dell'opera se noto

```
<titleStmt>
<title>Asclepius</title>
<author></author>
</titleStmt>
```

- indicazione dell'edizione a stampa di riferimento

```
<sourceDesc>
<p>Corpus Hermeticum, <hi rend="italic"> t. II, Traités
XII-XVIII</hi>, Asclepius, texte établi par A.D.Nock,
cinquième tirage revu, Paris 1992 (Collection des
Universités de France)</p>
</sourceDesc>
```

- indicazione delle eventuali correzioni materiali apportate al testo dell'edizione a stampa

```
<editorialDecl>
<p><hi rend="bold">Modifiche apportate al testo
dell'edizione di riferimento</hi></p>
<p>Si è corretto un errore di stampa al capitolo 34:
"impossible" sostituito con "impossibile"</p>
</editorialDecl>
```

³ Nei frammenti riportati qui sotto si trovano solo le indicazioni di struttura ma anche le indicazioni di formattazione, per esempio il tag `<hi rend="italic">`, perché come viene illustrato con maggiore dettaglio qui sotto, il file in formato TEI/XML è la fonte da cui tutte le altre versioni vengono ricavate.

- indicazione della licenza d'uso

```
<availability>
<p><ref
target="http://creativecommons.org/licenses/by-nc-
sa/3.0/it/" /></p>
</availability>
```

- descrizione essenziale della marcatura adottata

```
<editorialDecl>
<p><hi rend="bold">Note di trascrizione e codifica del
progetto digilibLT</hi></p>
<p>Abbiamo riportato solo il testo stabilito
dall'edizione di riferimento non seguendone
l'impaginazione. Sono stati invece esclusi dalla
digitalizzazione apparati, introduzioni, note e
commenti e ogni altro contenuto redatto da editori
moderni.</p>
<p> I testi sono stati sottoposti a doppia rilettura
integrale per garantire la massima correttezza di
trascrizione</p>
<p>Abbiamo normalizzato le U maiuscole in V seguendo
invece l'edizione di riferimento per la distinzione
u/v minuscole.</p>
<p> Non è stato normalizzato l'uso delle virgolette e
dei trattini.</p>
<p> Non è stata normalizzata la formattazione
nell'intero<hi rend="italic">corpus </hi>: grassetto,
corsivo, spaziatura espansa.</p>
<p>Sono stati normalizzati e marcati nel modo seguente
i diacritici.
<lb/>espunzioni: <![CDATA[<del>testo</del>]]> si
visualizza [testo]
<lb/>integrazioni:
<![CDATA[<supplied>testo</supplied>]] si visualizza
<testo>>
<lb/><hi rend="italic">loci desperati</hi>:
<![CDATA[<unclear>testo</unclear> o<unclear/> testo]]>
si visualizzano †testo† e †testo
```

```

<lb/> lacuna materiale:<![CDATA[<gap/>]]> si
visualizza [...]
<lb/> lacuna integrata dagli editori:
<![CDATA[<supplied><gap/></supplied>]] si
visualizza<...></p>
<p>Sono stati marcati i termini in lingua greca
<![CDATA[<foreign xml:lang="grc"> αγβ</foreign>]]></p>
<p>Ulteriori informazioni sulle norme di trascrizione
e codifica su <ref target="h
ttp://digiliblt.lett.unipmn.it/upload/docs/note_di_tra
scrizione_e_di_codifica.pdf">Note di trascrizione e
codifica di testi latini tardoantichi
[http://digiliblt.le
tt.unipmn.it/upload/docs/note_di_trascrizione_e_di_cod
ifica.pdf]</ref></p>
</editorialDecl>

```

- indicazione della o delle persone che hanno lavorato alla realizzazione dell'edizione digitale: dalla digitalizzazione, alla correzione, alla marcatura.

```

<respStmt>
<resp>Correzione linguistica</resp>
<name>Simona Rota</name>
<resp>Codifica XML</resp>
<name>Simona Musso</name>
</respStmt>
...
<publisher>digilibLT</publisher>
<pubPlace>Vercelli</pubPlace>
<date>2012</date>

```

Per alcune delle sezioni del TEI header il significato è ovvio o è già stato menzionato, come nel caso dell'indicazione della fonte dell'opera e delle correzioni apportate rispetto all'edizione di riferimento.

La descrizione essenziale della marcatura adottata ha lo scopo di rendere autosufficiente l'opera in formato digitale, il che si rende necessario dal momento che le opere possono essere liberamente scaricate dal sito della biblioteca perdendo così la connessione con la fonte delle informazioni che le riguardano.

L'indicazione dei nomi delle persone che hanno lavorato alla realizzazione dell'edizione digitale ha la motivazione fondamentale di far emergere i ruoli e le competenze di chi lavora al progetto senza avere un ruolo direttivo ma pure dando un contributo essenziale alla sua esistenza, e di valore scientifico indiscutibile; oltretutto di permettere a chiunque abbia lavorato alle edizioni della biblioteca (si tratta principalmente di Simona Musso con responsabilità anche di coordinamento, di Alice Borgna, Emanuela Ferroni, Laura Mosca, Nadia Rosso, Valentina Rinaldi, Beatrice Strona; ma non solo) di poter rendere conto del lavoro fatto nelle sedi opportune – quale ad esempio la VQR (valutazione della qualità della ricerca).

14.2.2. Distribuzione delle opere in molteplici formati

Le opere sono distribuite in cinque differenti formati:

- TEI/XML
- TXT
- PDF
- ePUB

ognuno dei quali risponde a scopi differenti.

Il formato TEI/XML, si potrebbe dire, raccoglie una parte della conoscenza che si possiede sul testo e che si intende descrivere e trasmettere in modo formalizzato. Infatti da questo formato vengono ricavati tutti gli altri con opportuni fogli di stile e scenari di trasformazione, allo scopo di evitare disallineamenti dello stato del testo (p. es. una correzione apportata al file in formato TXT ma non al file in formato TEI/XML). L'ambiente software utilizzato è Oxygen, e fogli di stile e scenari di trasformazione sono curati da Fabio Ciotti.

Il formato TXT è pensato primariamente per chi voglia fare ricerche sulle opere utilizzando programmi e metodologie che vanno dall'analisi testuale/produzione di concordanze fino all'analisi statistica multivariata, analyse des correspondences, e simili.

Il formato PDF è pensato soprattutto per chi non vuole rinunciare alla lettura e annotazione su carta (per quanto il formato permetta anche la lettura a schermo).

Il formato ePUB infine è concepito per rendere possibile la lettura, e/o studio, dei testi utilizzando un ambiente di lettura per ebook. Ambiente di lettura che può essere costituito primariamente da un

dispositivo dedicato (ebook reader), ma non si esclude un programma per lettura di ebook che gira su PC (come iBook per Mac oppure Calibre per Windows).

14.2.3. Licenza d'uso e accesso aperto

Tutta la biblioteca – sito e opere – è posto sotto licenza CC BY-NC-SA, cioè licenza Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo.



Fig. 14.2. Didascalia del piè di pagina presente in ogni pagina del sito della biblioteca digilibLT.

Sulle licenze Creative Commons (i cui siti di riferimento americano e italiano sono rispettivamente <http://creativecommons.org> e <http://creativecommons.it>) basti qui dire che esse, nate per il digitale, prevedono come assunto di base che gli oggetti digitali possano essere liberamente copiati e scambiati (mentre la tradizionale normativa sul diritto d'autore di fatto nega questa possibilità) e definiscono i vincoli sotto cui ciò può avvenire. I vincoli vengono scelti da chi produce i dati digitali (l'autore) e li mette a disposizione di altri. L'adozione delle licenze Creative Commons definisce in termini formali la scelta del progetto di rendere liberamente accessibili (accesso aperto) i prodotti della ricerca.

Nel caso del progetto digilibLT i vincoli sono:

- che venga menzionato l'autore (il progetto digilibLT con le persone che vi hanno lavorato): è il criterio "attribuzione";
- che non venga fatto un uso che comporti la produzione di un lucro economico: è il criterio "non commerciale"; esso non nega in termini di principio l'eventuale uso commerciale, che deve essere però concordato tra il soggetto che intende fare utilizzo commerciale e l'autore dell'opera;
- e infine che chi scarica i dati e li ridistribuisce li ridistribuisca senza cambiare la licenza d'uso: è il criterio "condividi allo stesso modo";

Chi vuole scaricare opere dal sito in uno dei formati descritti deve registrarsi fornendo alcuni dati elementari (nome e cognome, eventuale istituzione di appartenenza) ma soprattutto sottoscrivendo

il riconoscimento e l'accettazione della licenza d'uso dei testi, che vincola alla citazione della fonte e al mantenimento della licenza originaria in caso di redistribuzione dei contenuti.

14.2.4. Free software

Tutto il sito della biblioteca è costruito con free software cioè con software non proprietario ma disponibile al pubblico: dal sistema operativo Ubuntu, al server http Apache, ai linguaggi Perl e MySQL, al gestore della base di dati XTF, fino alle singole utilities (per esempio quella che controlla con un captcha l'accesso degli utenti al sito). Per una scelta di principio, cioè che la ricerca realizzata con fondi pubblici utilizza strumenti pubblicamente disponibili, ma anche per un motivo pratico: che una volta terminato il finanziamento del progetto il sito della biblioteca possa continuare a funzionare senza costi vivi per le licenze. Non si è scelto di sviluppare (di far sviluppare) software specifico, dunque, ma di usare componenti preesistenti, perché lo scopo primario era di costruire la biblioteca, di renderla funzionante il più rapidamente possibile, non altro.

14.2.5. Ambiente di studio

Quando si seleziona un'opera a cui si vuole accedere, per prima cosa si giunge ad una pagina specifica dell'opera, che contiene una concisa scheda critica storico-letteraria di presentazione scritta appositamente per la biblioteca dai latinisti del progetto e – soprattutto – contiene una serie di riferimenti bibliografici specifici per l'opera stessa. Lo scopo è di inserire le opere in un contesto di studio e di ricerca che risulti segnali ciò che di più valido e recente è stato pubblicato.

14.2.6. Nascita dal basso

La biblioteca nasce non ad opera di uno o più soggetti dell'ambito bibliotecario o ICT, ma ad opera essenzialmente di un gruppo di latinisti e filologi, che possiedono in misura variabile competenze nell'ambito digitale. E nasce proprio perché essi individuano una lacuna nel loro ambito di studi: la mancanza di opere affidabili in formato digitale. Essi pertanto si sono messi all'opera per costruire lo strumento di cui avevano bisogno (la biblioteca) definendone le condizioni di esistenza, se così si può dire:

- qualità delle edizioni,
- accesso aperto,
- free software,
- download libero,

insomma le caratteristiche qui sopra concisamente elencate e descritte.

14.2.7. digilibLT e (le) altre biblioteche digitali del latino

Prima di digilibLT una biblioteca digitale del latino tardoantico propriamente detta non esisteva. Parte dei testi latini tardi sono disponibili in importanti biblioteche digitali quali

- BIA, Bibliotheca Iuris Antiqui, diretta da Nicola Palazzolo, che raccoglie i testi giuridici;
- Musisque deoque, diretta da Paolo Mastandrea, che raccoglie la poesia latina dalle origini fino al Rinascimento;
- i Grammatici Latini raccolti e curati da Nino Marinone all'Università di Torino;

e se si allarga la visione ai testi medievali – per i quali peraltro in quanto testi cristiani esistono importanti raccolte specifiche – sono da ricordare almeno

- ALIM – Archivio della latinità italiana medievale – diretto da Gian Carlo Alessio;
- il Corpus Rithmorum Musicum diretto da Francesco Stella.

Vari contatti e progetti di collaborazione esistono o sono stati tentati con questi progetti, soprattutto nella prospettiva di una 'via italiana alla biblioteca digitale globale del latino' se così si può dire, dal momento che in Italia hanno sede biblioteche digitali il cui contenuto costituisce una parte veramente rilevante della produzione in lingua latina dalle origini al medioevo. In qualche caso però l'adozione di sistemi proprietari per la gestione della biblioteca digitale ha reso impossibile la collaborazione. Un aspetto importante che facilita e permette di progettare l'integrazione è dato dal fatto che molte di queste biblioteche utilizzano marcatori TEI per descrivere le caratteristiche dei testi. Laddove non via siano codifica in standard TEI o almeno architetture aperte la collaborazione risulta impossibile indipendentemente dalle intenzioni degli studiosi.

Contatti sono in corso da tempo per rendere disponibili i testi di digilibLT anche attraverso Internet Culturale ed Europeana.

14.3. Prospettive future per lo sviluppo di digilibLT

L'orizzonte futuro è quello che è stato definito il "global latin corpus", la "biblioteca digitale del latino". Senza proporsi mete alte e ambiziose come quelle di Gregory Crane, recente Humboldt Professor a Lipsia con l' Open Philology Project, di cui ha scritto

The long term goal of the Open Greek and Latin Project is to represent the full surviving corpus of Greek and Latin sources, including transcriptions from every print source, this will include not only print books but also manuscripts, inscriptions, ostraca, papyri, vases, etc. and will cover the full range of Greek and Latin sources, from the Homeric Epics through post-classical Greek and Latin to the present." (Crane 2013)

L'obiettivo del gruppo di ricerca di digilibLT è per intanto di integrare i testi latini tardoantichi di digilibLT con i testi già presenti nel PHI cdrom in quanto tale raccolta, a norma della legge italiana e delle leggi europee sul diritto d'autore, è ormai libera da diritti, e con i Grammatici Latini. Con ulteriori ampliamenti che potrebbero riguardare gli scritti giuridici e la poesia, ma anche autori cristiani e medievali, se la "biblioteca digitale del latino" si costituisce. Contatti e iniziative sono in corso in varie forme, tutto è ancora da definire in dettaglio, ma l'orizzonte è quello.

14.4. ePUB, ebook, e gli usi di una biblioteca di testi latini

ePUB, come è noto, è il formato aperto per ebook, basato su XML.

Rendere disponibili le opere latine tarde in formato ePUB significa prefigurare che uno studioso possa costruirsi una biblioteca personale e portatile (*facilmente portatile*) e che arrivi ad *usare intensamente le funzioni di ricerca* all'interno della collezione, funzioni offerte dal dispositivo (se si lavora con un ebook reader) o dal software (se si lavora con un computer tradizionale) che dà accesso alla biblioteca. Uno strumento di consultazione che svolge la funzione prima associata ad una serie di operazioni fisiche: alzarsi dal tavolo, individuare e prendere in mano il

volume che contiene l'opera, aprirlo, sfogiarlo e iniziare la ricerca del passo che interessa! E ripetere quest'ultima operazione tante volte quanti sono i passi che si vogliono reperire. Questa descrizione astratta parte da due presupposti che non sempre si verificano: in primo luogo, che lo studioso possieda l'intera raccolta delle opere della letteratura in cui si collocano i suoi interessi; in secondo luogo che lo studioso conosca gli estremi del passo che vuole citare. Pochi studiosi possiedono una biblioteca ragionevolmente completa della letteratura che studiano, in genere ciò che possiedono è centrato su un periodo, o un tema di interesse dello studioso. Spesso è necessario ricorrere a prestito o consultazione in biblioteca. E i passi spesso sono ricordati per il loro contenuto, non per la loro posizione nell'opera.

Qui invece si prospetta (*si prospetta*, perché il progetto digilibLT sta *iniziando* a mettere a disposizione questo tipo di raccolta, con un orientamento netto e chiaro verso questo scopo) una situazione in cui lo studioso possa avere a disposizione l'intera biblioteca della lingua e letteratura che studia, insieme con funzioni di ricerca e di selezione dei passi. Già questo solo basta delineare i tratti di un'evoluzione nello studio dei testi. Come tutte le evoluzioni essa avverrà lentamente, all'inizio non è (non sarà) nemmeno un'evoluzione ma solo uno sporadico manifestarsi di scelte, operazioni, comportamenti, differenti da quelli dominanti.

Il possesso – non bibliofilo, ma funzionale allo studio – delle edizioni delle opere che si studiano (e che si amano! si pensi al francese "apprendre par coeur") è una componente ineliminabile dell'atteggiamento dello studioso, basta ricordare nella sua esemplarità la vicenda di Petrarca che nel 1352, partito dalla Provenza verso l'Italia, interrompe il viaggio e torna indietro perché il maltempo rischiava di rovinare la biblioteca personale che portava con sé. Il senso dell'avere le opere con sé rimanda alla conoscenza delle opere come base per il loro studio. DigilibLT prospetta una situazione in cui ogni studioso di latino potrà possedere tutte le opere della letteratura, senza difficoltà di conservazione.

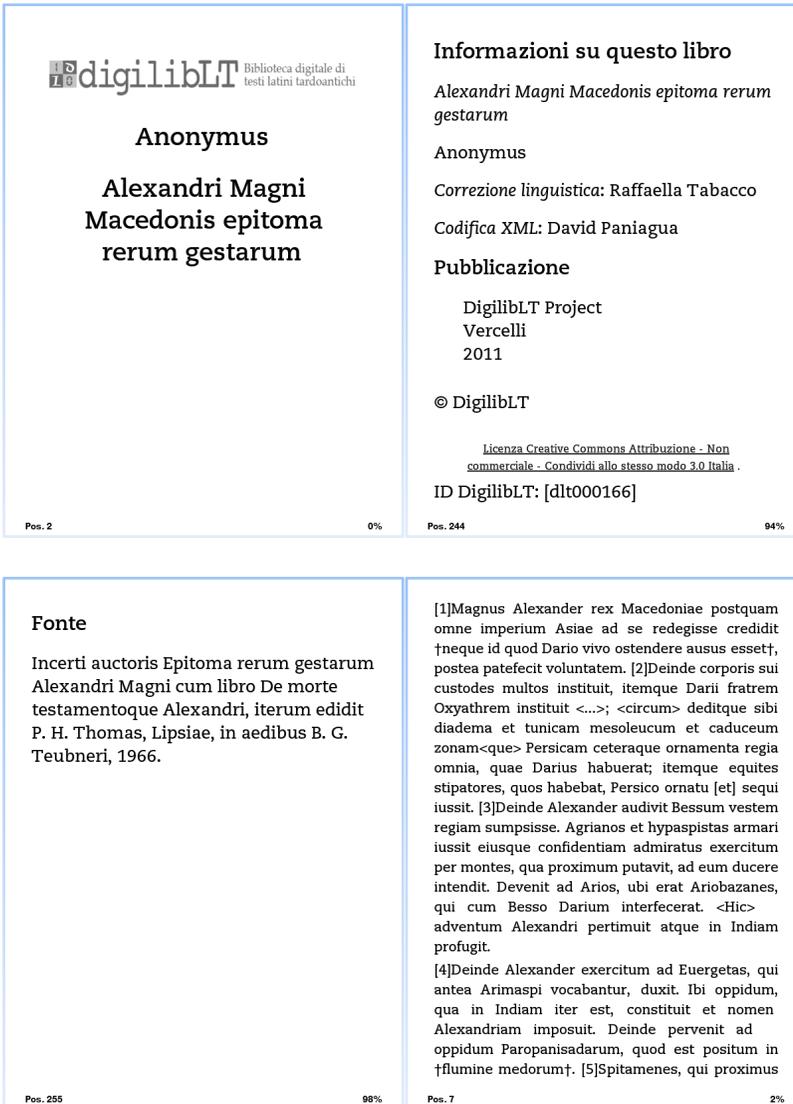


Fig. 14.3. Pagina da un'opera in formato ePUB disponibile nella biblioteca digilibLT.

Ciò inevitabilmente amplierà le prospettive di studio, anche perché si darà una risposta ad uno dei problemi tipici dello studio basato sulla carta: il reperimento dei passi rimane un problema sempre 'scoperto', privo di soluzioni efficaci. Infatti la lettura e la conoscenza approfondita, per quanto accompagnate da schedatura, non aiutano quando nell'elaborazione mentale e interiore dei testi letti si manifesta

l'importanza di un passo non schedato, quando si ricorda di aver letto un passo che serve per lo sviluppo del pensiero e dell'argomentazione che non è stato registrato. E la schedatura ha il suo punto debole nel presupposto razionalistico che sia importante ciò che appare tale nel momento della lettura; mentre proprio l'affiorare alla coscienza dell'importanza di un passo non schedato mostra (senza addentrarsi in discorsi complessi sulle modalità dello sviluppo e dell'evoluzione del pensiero in relazione alla 'scoperta scientifica') che la lettura diretta dei testi innesca processi non coscienti di elaborazione dei testi e del loro contenuto e che tale lettura è dunque indispensabile ed ineliminabile.

Acquista rilevanza dunque l'operazione concettualmente semplice del reperimento di un passo che nel ricordo potrebbe anche presentarsi non del tutto preciso proprio a causa del processo mentale che lo porta in evidenza: imprecisioni lessicali dovute a sinonimie vere o presunte, abbreviazioni, sostituzioni. Giunge in aiuto, ad integrazione e potenziamento dell'intelletto umano (non in sostituzione), uno strumento digitale: la raccolta completa delle opere integrata da uno strumento di ricerca che permette di esplorare i testi cercando la corrispondenza tra ciò che si ricorda e ciò che i testi effettivamente contengono. Una protesi per la mente. Nessuno vede negli occhiali che agevolano o addirittura consentono la lettura una protesi che depriva l'individuo delle sue capacità intellettuali; e qui è lo stesso, il programma di gestione di una libreria di ebook, con le sue funzioni di ricerca è una protesi che aiuta nella lettura dei testi al livello elementare delle parole ivi contenute.

Questo conforta e rassicura che sia preoccupato temendo che l'entrata del digitale nello studio dei testi snaturi lo studio stesso, ne muti radicalmente le caratteristiche con cui dalla tradizione esso ci è stato trasmesso. Se poi qualcuno pensasse che non ci sia bisogno di opere in formato ePUB perché c'è già il PDF, non si devono confondere PDF ed ePUB, che nascono con scopi differenti anche se molteplici malintesi vedono nel PDF un formato per ebook. In realtà le pagine PDF sono concepite per una visualizzazione rigida, in PDF il testo che scorre e si adatta all'ampiezza dei margini, teoricamente possibile, in realtà si incontra raramente. Per non dire delle difficoltà che si incontra se si vuole effettuare una ricerca su un'intera collezione e non solo all'interno di una singola opera.

Quanto fin qui detto rimanda al dibattito in corso in blog e siti sul tema del testo come "fonte di dati" e sui modi per sfruttarla efficacemente. La ricerca delle parole come modo per individuare le opere rilevanti per un certo discorso è la forma elementare del riconoscimento del testo come fonte di dati, e una valida presentazione di questo approccio è di Hugh McGuire in (McGuire 2013). Dalla ricerca dei passi in base al testo che essi contengono, è breve il passo verso uno studio dei testi strettamente connesso con la ricerca di informazioni veicolate dalle parole.

14.5. Conclusioni

Le conclusioni sono duplici. Da una parte, come scrive Franco Moretti,

In the last few years, literary studies have experienced what we could call the rise of quantitative evidence. This had happened before of course, without producing lasting effects, but this time it's probably going to be different, because this time we have digital databases, and automated data retrieval. [...] today, we can replicate in a few minutes investigations that took a giant like Leo Spitzer months and years of work. When it comes to phenomena of language and style, we can do things that previous generations could only dream of. (Moretti 2011).

Perché ciò possa avvenire servono i testi in formato digitale, accessibili in "digital databases" come scrive Moretti, ma meglio ancora se distribuiti da biblioteche digitali: la differenza sta nel maggiore o minore senso di cura per il testo nell'arco della sua esistenza, e di comunità degli studi centrata sui testi.

Dall'altra, la biblioteca digitale di cui si sono stati presentati alcuni aspetti caratterizzanti come si è visto è chiaramente orientata su molteplici linee fortemente qualificanti: accesso aperto, free software, cura filologica nella scelta delle edizioni, adozione di standard aperti, ... L'esperienza della sua realizzazione, innovativa rispetto all'esistente per questi molteplici aspetti, è a disposizione della comunità degli studiosi e di chi ne volesse replicare o sviluppare il modello.

14.6. Bibliografia

- BARDINI TH. (2000). *Bootstrapping: Douglas Engelbart, coevolution, and the origins of personal computing*, Stanford University Press.
- CRANE G. (2013). *Possible Jobs in Digital Humanities at Leipzig*, URL=<http://sites.tufts.edu/perseusupdates/2013/02/14/possible-jobs-in-digital-humanities-at-leipzig/> [ultima visita 18.2.2013].
- MCGUIRE H. (2013). *A Publisher's Job Is to Provide a Good API for Books*, URL=<http://toc.oreilly.com/2013/02/a-publishers-job-is-to-provide-a-good-api-for-books.html> [ultima visita 18.2.2013].
- MICHEL J-B. *et al.* (2010), *Quantitative Analysis of Culture Using Millions of Digitized Books*, «Science», December 2010, URL= <https://www.sciencemag.org/content/331/6014/176.full> [ultima visita 19.2.2013].
- MORETTI F. (2011). *Network theory and Plot analysis*, URL=<http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> [ultima visita 20.2.2013].
- SALARELLI A., TAMMARO A.M. (2006). *La biblioteca digitale. Nuova edizione interamente riveduta e ampliata*, Editrice Bibliografica.
- WITTEN I.H., BAINBRIDGE D., NICHOLS D.M. (2009). *How to Build a Digital Library*, Morgan Kaufmann.

15. L'ontologia BIA-Net: una base per la ricerca di informazioni secondo rapporti di rilevanza nella Bibliotheca Iuris Antiqui

Marco Giunti¹, Giuseppe Sergioli², Giuliano Vivonet³

Abstract. *BIA-Net* è la versione accessibile via Web del vasto archivio della *Bibliotheca Iuris Antiqui*, che consta di tre grandi parti: *Opera*, un archivio bibliografico di oltre 50.000 opere di diritto romano, *Thesaurus*, un sistema di classificazione dei documenti presenti in *Opera*, e infine *Fontes*, l'archivio digitalizzato delle fonti dei diritti dell'antichità. Questo lavoro presenta una proposta di ontologia per *BIA* (*Ontologia BIA-Net*), mediante la descrizione dettagliata delle classi e relazioni costitutive di ciascuno dei suoi tre archivi e delle relazioni fra gli archivi stessi. Nella parte finale vengono anche proposte ulteriori prospettive di ricerca, inerenti allo sviluppo di due applicazioni di tipo semantico, un motore di ricerca e un browser, tramite cui rendere la base di conoscenze di *BIA* semanticamente interrogabile e navigabile.

Parole chiave: *BIA*, *Bibliotheca Iuris Antiqui*, ontologia, rete semantica, grafo RDF, OWL, Web semantico, browser semantico, motore di ricerca semantico, rilevanza, pertinenza.

¹ ALOPHIS, Università di Cagliari, Italy. e-mail: giunti@unica.it.

² ALOPHIS, Università di Cagliari, Italy. e-mail: giuseppe.sergioli@gmail.com.

³ ALOPHIS, Università di Cagliari, Italy. e-mail: giuliano.vivanet@unica.it.

15.1. Introduzione

Il progetto *BIA-Net: strumenti semantici e applicazioni multimediali per la ricerca nella Bibliotheca Iuris Antiqui*, finanziato nell'ambito del PRIN 2009, nasce dall'esigenza di rendere accessibile sul Web il patrimonio scientifico della *Bibliotheca Iuris Antiqui* (*BIA*). Quest'ultimo è un sistema informativo integrato sui diritti dell'antichità, già in precedenza digitalizzato, ma finora accessibile solo mediante CD-ROM (Palazzolo 2002). Il progetto vede la partecipazione di cinque unità afferenti all'Università di Catania, Università di Perugia, Consiglio Nazionale delle Ricerche, Università 'Magna Graecia' di Catanzaro e Università di Cagliari.

Obiettivo finale è la realizzazione di un prototipo di biblioteca digitale in rete, *BIA-Net*, utilizzando tecnologie open source e standard internazionali per la gestione degli archivi e dell'interfaccia Web, oltre che i linguaggi del Web Semantico (Berners Lee et al. 2001; Antoniou e van Harmelen 2008) per la rappresentazione e codifica della base di conoscenze di *BIA*.

L'archivio di *BIA* è composto da tre parti fondamentali: *Opera*, un archivio bibliografico di oltre 50.000 opere che trattano il diritto romano e, più in generale, i diritti dell'antichità; *Thesaurus*, un sistema di classificazione dei documenti presenti in *Opera*, costituito da una gerarchia di alcune centinaia di classi e da circa 10.000 concetti che descrivono tali classi; e infine *Fontes*, l'archivio digitalizzato delle fonti dei diritti dell'antichità.

Nel quadro degli obiettivi generali del progetto, l'unità di Cagliari offre il proprio supporto teorico a una revisione del *Thesaurus* di *BIA* (Palazzolo 1995) ai fini della realizzazione di un sistema di reperimento automatico dell'informazione secondo rapporti di rilevanza. In questa prospettiva, è fondamentale definire l'ontologia sottostante ai tre archivi di cui *BIA* è composta. L'utilità di questo tipo di rappresentazione consiste nell'esplicitazione della *struttura reticolare* delle informazioni contenute nei tre archivi di *BIA*. Tale struttura potrà così essere utilizzata da applicazioni capaci di reperire automaticamente le informazioni più rilevanti per una data informazione di partenza *c*. A seconda del modo in cui si interpreti l'informazione data *c*, si potranno sviluppare applicazioni di tipi diversi: (i) un motore di ricerca semantico, quando *c* sia intesa come la chiave per una *richiesta*

(query), oppure (ii) un browser semantico, quando *c* sia intesa come il punto di partenza di una vera e propria *esplorazione* della struttura reticolare stessa.

In questo contributo, presentiamo una proposta di ontologia per *BIA* (*Ontologia BIA-Net*), fornendo una descrizione delle classi e relazioni costitutive di ciascuno dei suoi tre archivi, e delle relazioni fra gli archivi stessi. L'ontologia è stata implementata in Web Ontology Language Description Logic – OWL DL (W3C 2004a), mediante Protégé 3.4.7⁴. Allo stato attuale, l'ontologia comprende soltanto lo *schema* ontologico costituito da classi e relazioni (parte intensionale), che dovrà poi essere utilizzato per il *popolamento* dell'ontologia (parte estensionale). Solo dopo che il popolamento sarà realizzato, si potranno sviluppare le specifiche applicazioni di tipo semantico (un motore di ricerca e un browser), volte al miglioramento della ricerca e della fruibilità delle informazioni contenute nei tre archivi di *BIA*.

15.2. L'Ontologia BIA-Net

Come anticipato, l'archivio di *BIA* è costituito da tre parti interconnesse: *Opera*, *Thesaurus* e *Fontes*. *Opera* è un archivio bibliografico di oltre 50.000 opere che trattano del diritto romano e, più in generale, dei diritti dell'antichità. Nel *Thesaurus* si trovano delle *classi*, che permettono di classificare i contenuti dei documenti presenti in *Opera*, e dei *concetti* che descrivono tali classi. Infine, *Fontes* è l'archivio delle fonti dei diritti dell'antichità. Si fornisce di seguito una dettagliata descrizione dei tre archivi, attraverso l'esplicitazione delle classi e della relazioni che ne costituiscono l'ontologia.

Si tenga presente che una qualsiasi relazione il cui nome è del tipo 'X1aaaaX2...Xn' è implementata come una Object property di OWL, e dunque ha inversa. Tuttavia, per semplicità, l'inversa non è indicata né nelle figure né nella successiva tabella. Tutte le altre relazioni sono Datatype properties di OWL; i loro valori sono dunque literal e nessuna di tali relazioni ha inversa.

⁴ The Protégé Ontology Editor and Knowledge Acquisition System:
<http://protege.stanford.edu/>.

15.2.1. Thesaurus

Il *Thesaurus* si presenta in due forme diverse ma a tutti gli effetti equivalenti: classificato o alfabetico. Facendo riferimento alla forma classificata, ciascuna classe del *Thesaurus* è identificata da un *codice numerico*. A ciascuna classe corrisponde una *vedetta* che si può presentare tradotta in cinque lingue (italiano, inglese, tedesco, francese, spagnolo), oppure solo in latino. Per esempio, la vedetta della classe con codice '1' è espressa dai cinque termini: 'FONTI', 'SOURCES', 'QUELLEN', 'SOURCES', 'FUENTES', mentre la classe il cui codice è la stringa '1.1.1.1' ha come vedetta solo il termine latino 'LEGES'. Si noti che a classi diverse può corrispondere la stessa vedetta. Per esempio, le classi 1.1.1.1 e 1.1.2.5.1 hanno la stessa vedetta LEGES.

Ciascun codice rappresenta biunivocamente una classe del *Thesaurus*. Le classi possono essere fra loro nella relazione di *sottoclasse* (in questo caso i codici sono del tipo: '1'; '1.1'; '1.1.1'). Classi appartenenti allo stesso livello gerarchico hanno codici del tipo '1.1'; '1.2'; '1.3', e così via.

Inoltre, ciascuna classe può avere associati dei *descrittori*, che sono concetti semanticamente pertinenti con la stessa. Essi sono indicati nel *Thesaurus* classificato mediante termini *latini*. Per esempio, i descrittori della classe 1.1.1.3 sono dieci concetti corrispondenti ai seguenti termini latini o latinizzati: 'basilica', 'de actionibus', 'ecloga', 'epanogoge aucta', 'manuale legum', 'prochiron', 'synopsys basilicorum', 'synopsis minor', 'Theophili paraphrasis', 'Tipucitus'. Si noti che lo stesso concetto può essere descrittore di più classi. Per esempio, il concetto aurum è descrittore di ambedue le classi 3.2.2 e 5.7.

A ciascun descrittore di una data classe può risultare associato un gruppo di altri concetti, espressi da termini latini. Essi, al solo fine della descrizione della classe in questione, sono *sussunti* nel relativo descrittore. Il rapporto tra ciascuno di tali concetti, il descrittore a cui esso è associato e la classe descritta determina quindi una relazione ternaria, che denominiamo 'X1èSussuntoInX2è-DescrittoreDiX3'.

Tra i concetti sussunti da un dato descrittore di una data classe vi può essere una relazione di equivalenza che può esprimere sinonimie, varianti ortografiche, o simili. Tale equivalenza non è da intendersi in modo assoluto, ma soltanto relativamente al particolare descrittore sotto il quale i concetti sono sussunti e alla particolare classe che tale descrittore

descrive. Il rapporto tra i due concetti equivalenti, il descrittore che li sussume e la classe descritta determina quindi una relazione quaternaria, che denominiamo 'X1equivalenteX2èSussuntoInX3èDescrittoreDiX4'.

Analoga relazione di equivalenza può sussistere anche tra i concetti descrittori di una classe. In tal caso, il rapporto tra i due descrittori equivalenti e la classe descritta determina una relazione ternaria, che denominiamo 'X1equivalenteX2èDescrittoreDiX3'.

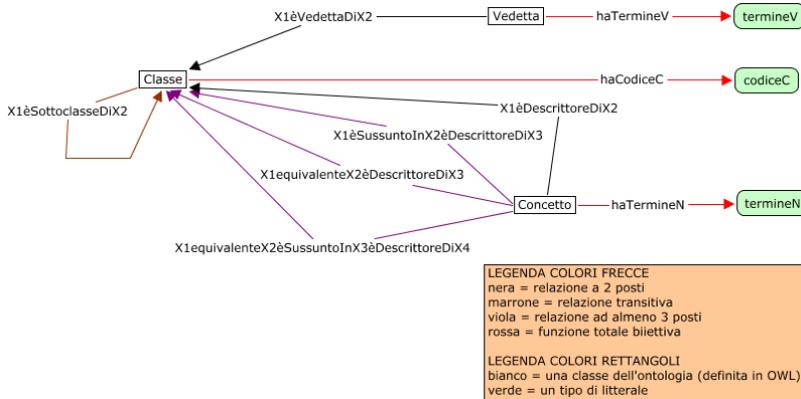


Fig. 15.1. Schema dell'ontologia del *Thesaurus*.

15.2.1.1 Classi dell'ontologia del *Thesaurus*

- Col rettangolo denominato 'Classe' (si veda la fig. 15.1) si denota la classe dell'ontologia che comprende tutte le classi del *Thesaurus* di *BIA*. Si tenga sempre presente che le classi di *BIA* non sono trattate come classi dell'ontologia, ma come individui⁵.

Relazioni con dominio *Classe*

1. *Classe* è il dominio della relazione *haCodiceC*, il cui codominio è formato da tutti i literali che sono codici di una classe di *BIA*. In fig. 15.1 questo tipo di letterale è indicato da un rettangolo verde denominato 'codiceC'.
2. Le classi di *BIA* possono essere in relazione tra loro in un rapporto

⁵ Se le classi di *BIA* fossero trattate come classi dell'ontologia (e quindi la relazione *X1èSottoclasseDiX2* fosse identificata con la relazione *rdfs:subClassOf*, che è una primitiva di OWL) l'ontologia stessa non sarebbe più esprimibile in OWL DL, ma solo in OWL Full.

di inclusione; tale relazione è indicata da 'X1èSottoclasseDiX2'. La relazione X1èSottoclasseDiX2 deve essere intesa come interna alla classe delle classi di *BIA*. In altri termini, sia il dominio che il codominio di tale relazione è Classe, ovvero la classe di tutte le classi del *Thesaurus* di *BIA*. Inoltre, la relazione X1èSottoclasseDiX2 è transitiva, e la sua inversa è la relazione X2haSottoclasseX1.

- Col rettangolo denominato 'Vedetta' si denota la classe dell'ontologia che comprende tutte le vedette del *Thesaurus* di *BIA*.

Relazioni con dominio Vedetta

1. Vedetta è il dominio della relazione haTermineV, il cui codominio è formato dai nomi di tutte le vedette di *BIA*. Nella fig. 15.1 questo tipo di litterale è indicato da un rettangolo verde denominato 'termineV'⁶.
 2. La relazione X1èVedettaDiX2 ha dominio Vedetta e codominio Classe. La relazione inversa è X2haVedettaX1.
- Col rettangolo denominato 'Concetto' si denota la classe dell'ontologia che comprende tutti i concetti del *Thesaurus* di *BIA*.

Relazioni con dominio Concetto

1. Concetto è il dominio della relazione haTermineN, il cui codominio è formato dai nomi di tutti i concetti di *BIA*. In fig. 15.1 questo tipo di litterale è indicato da un rettangolo verde denominato 'termineN'.
2. La relazione X1èDescrittoreDiX2 ha dominio Concetto e codominio Classe. La relazione inversa è X2haDescrittoreX1.
3. La relazione X1èSussuntoInX2èDescrittoreDiX3 è una relazione ternaria con primo dominio Concetto, secondo dominio Concetto e terzo dominio Classe.

⁶ In realtà siccome abbiamo visto che le vedette possono essere espresse in 6 lingue diverse (latino, italiano, inglese, tedesco, francese, spagnolo), si deve introdurre una relazione diversa per ciascuna lingua. Tali relazioni sono: haTermineVlat, haTermineVit, haTermineVde, haTermineVfr, haTermineVes. Per semplicità, in fig. 15.1 è mostrata soltanto la generica relazione haTermineV.

4. La relazione $X1 \text{equivalente} X2 \text{è} \text{SussuntoIn} X3 \text{è} \text{DescrittoreDi} X4$ è una relazione quaternaria con primo dominio Concetto, secondo dominio Concetto, terzo dominio Concetto e quarto dominio Classe.
5. La relazione $X1 \text{equivalente} X2 \text{è} \text{DescrittoreDi} X3$ è una relazione ternaria con primo dominio Concetto, secondo dominio Concetto e terzo dominio Classe.

15.2.2. Opera

Per ogni opera sono indicati il titolo, l'autore, la pubblicazione (rivista, editore, atti di, ecc.), il numero delle pagine (per i libri) o i numeri della prima e ultima pagina (per gli articoli), l'anno di pubblicazione, e infine il testo (che può essere o l'abstract dell'opera – se l'opera è un articolo, o l'indice – se l'opera è un libro).

15.2.2.1 Classi dell'ontologia di Opera

- Facendo riferimento alla fig. 15.2, col rettangolo denominato 'Autore' si denota la classe dell'ontologia che comprende tutti gli autori delle opere dell'archivio *Opera* di *BIA*.

Relazioni con dominio Autore

1. Autore è il dominio della relazione $haNomeA$, il cui codominio è formato dai nomi di tutti gli autori delle opere dell'archivio *Opera* di *BIA*. In fig. 15.2 questo tipo di litterale è indicato da un rettangolo verde denominato 'nomeA'.
- Col rettangolo denominato 'Pubblicazione' si denota la classe dell'ontologia che comprende tutte le pubblicazioni (riviste, editori, atti di convegni, ecc.) dell'archivio *Opera* di *BIA*.

Relazioni con dominio Pubblicazione

1. Pubblicazione è il dominio della relazione $haNomeP$, il cui codominio è formato dai nomi di tutte le pubblicazioni dell'archivio *Opera* di *BIA*. In fig. 15.2 questo tipo di litterale è indicato da un rettangolo verde denominato 'nomeP'.

- Col rettangolo denominato 'Anno' si denota la classe dell'ontologia che comprende tutti gli anni.

Relazioni con dominio Anno

1. Anno è il dominio della relazione haAnno_aCdC, il cui codominio è formato da tutti i literali che esprimono gli anni calcolati secondo il calendario gregoriano in una forma standard (per esempio, secondo il tipo xsd:gYear di XML Schema 1.0). In fig. 15.2 questo tipo di literal è indicato da un rettangolo verde denominato 'anno_aCdC'.

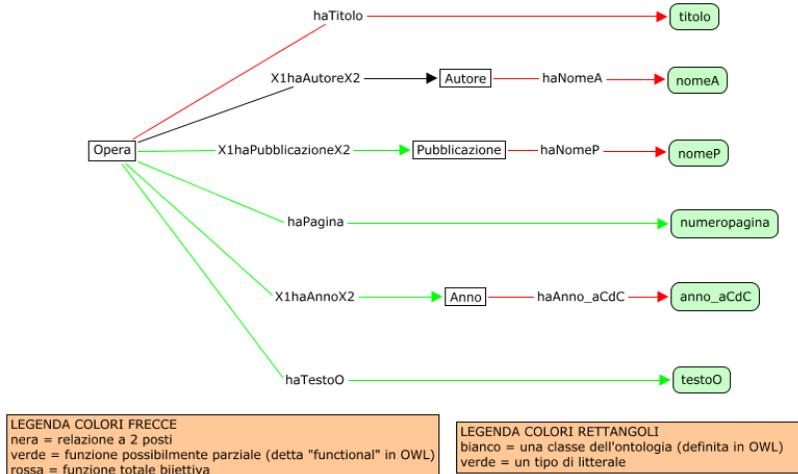


Fig. 15.2. Schema dell'ontologia di *Opera*.

- Col rettangolo denominato 'Opera' si denota la classe dell'ontologia che comprende tutte le opere dell'archivio *Opera* di *BIA*.

Relazioni con dominio Opera

1. *Opera* è il dominio della relazione haTitolo, il cui codominio è formato dai titoli di tutte le opere dell'archivio *Opera* di *BIA*. In fig. 15.2 questo tipo di literal è indicato da un rettangolo verde denominato 'titolo'.
2. *Opera* è il dominio della relazione haPagina, il cui codominio è formato dai literali che esprimono i numeri di pagina di tutte le opere dell'archivio *Opera* di *BIA*. In fig. 15.2 questo tipo di literal è indicato da un rettangolo verde denominato 'numeropagina'.

3. Opera è il dominio della relazione haTestoO, il cui codominio è formato da tutti i testi (l'abstract dell'opera – se l'opera è un articolo, o l'indice – se l'opera è un libro) di opere contenuti nell'archivio *Opera* di *BIA*. In fig. 15.2 questo tipo di letterale è indicato da un rettangolo verde denominato 'testoO';
4. La relazione X1haAutoreX2 ha dominio Opera e codominio Autore. La relazione inversa è X2èAutoreDiX1.
5. La relazione X1haPubblicazioneX2 ha dominio Opera e codominio Pubblicazione. La relazione inversa è X2èPubblicazioneDiX1.
6. La relazione X1haAnnoX2 ha dominio Opera e codominio Anno. La relazione inversa è X2èAnnoDiX1.

15.2.3. Fontes

Ciascuna *fonte* è univocamente identificata da un *codice*. Inoltre, ciascuna fonte è suddivisa in parti elementari, dette *luoghi fonte*, che sono anch'essi univocamente identificati da *codici*. I luoghi fonte possono essere o di tipo *testo* (in tal caso potremmo trovare, oltre al *testo vero e proprio*, anche una *subscriptio*) o di tipo *rubrica* (in tal caso avremo il *testo della rubrica*). Inoltre, alcuni luoghi fonte (sia di tipo testo, che di tipo rubrica) possono contenere una *inscriptio*.

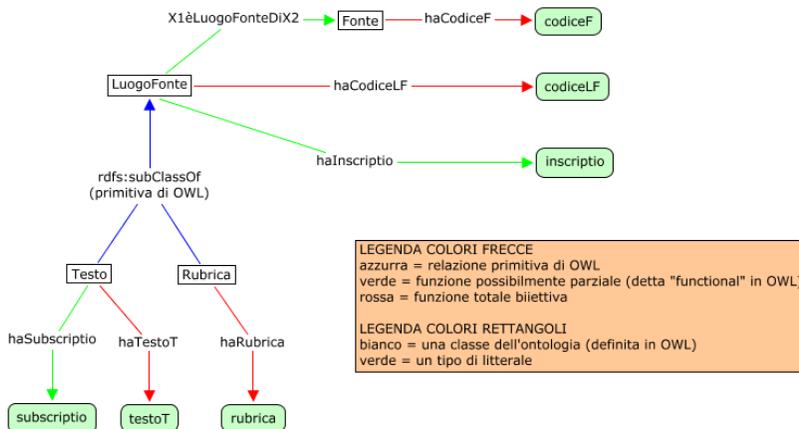


Fig. 15.3. Schema dell'ontologia di *Fontes*.

15.2.3.1. Classi dell'ontologia di Fontes

- Facendo riferimento alla fig. 15.3, col rettangolo denominato 'Fonte' si denota la classe dell'ontologia che comprende le fonti dei diritti dell'antichità.

Relazioni con dominio Fonte

1. Fonte è il dominio della relazione haCodiceF, il cui codominio consiste di tutti i codici delle fonti. In fig. 15.3 questo tipo di litterale è indicato da un rettangolo verde denominato 'codiceF'.
- Col rettangolo denominato 'LuogoFonte' si denota la classe dell'ontologia che comprende le unità elementari (luoghi fonte) di cui si compongono le fonti.

Relazioni con dominio LuogoFonte

1. LuogoFonte è il dominio della relazione haCodiceLF, il cui codominio consiste di tutti i codici dei luoghi fonte. In fig. 15.3 questo tipo di litterale è indicato da un rettangolo verde denominato 'codiceLF'.
 2. LuogoFonte è il dominio della relazione haInscriptio, il cui codominio è formato da tutti i litterali che costituiscono l'eventuale inscriptio del luogo fonte. In fig. 15.3 questo tipo di litterale è indicato da un rettangolo verde denominato 'inscriptio'.
 3. La relazione $X1 \text{ è } \text{LuogoFonte} \text{ Di } X2$ ha dominio LuogoFonte e codominio Fonte. La relazione inversa è $X2 \text{ ha } \text{LuogoFonte} X1$.
- Col rettangolo denominato 'Rubrica' si denota la classe dell'ontologia che comprende tutti i luoghi fonte di tipo rubrica. La classe Rubrica è quindi una sottoclasse della classe LuogoFonte.

Relazioni con dominio Rubrica

1. Rubrica è il dominio della relazione haTestoR, il cui codominio è formato da tutti i litterali che costituiscono le rubriche stesse. In fig. 15.3 questo tipo di litterale è indicato da un rettangolo verde denominato 'rubrica'.

- Col rettangolo denominato 'Testo' si denota la classe dell'ontologia che comprende tutti i luoghi fonte di tipo testo. La classe Testo è quindi una sottoclasse della classe LuogoFonte⁷.

Relazioni con dominio Testo

1. Testo è il dominio della relazione ha TestoR, il cui codominio è formato da tutti i literal che costituiscono il vero e proprio testo del luogo fonte. In fig. 15.3 questo tipo di literal è indicato da un rettangolo verde denominato 'testoT'.
2. Testo è il dominio della relazione haSubscriptio, il cui codominio è formato da tutti i literal che costituiscono l'eventuale subscriptio di un testo. In fig. 15.3 questo tipo di literal è indicato da un rettangolo verde denominato 'subscriptio'.

15.2.4. Relazioni fra i tre archivi

Oltre alle relazioni interne a ciascuno dei tre archivi *Thesaurus*, *Opera* e *Fontes*, sono presenti anche relazioni che li interconnettono.

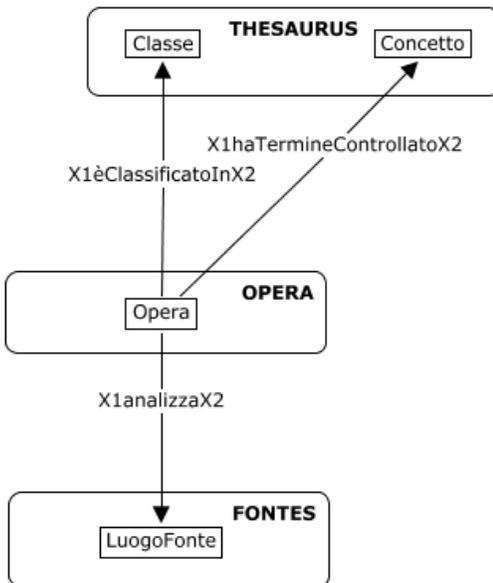


Fig. 15.4. Relazioni fra i tre archivi di BIA.

⁷ Testo e Rubrica costituiscono una partizione della classe LuogoFonte (ovvero, sono l'una il complemento dell'altra).

Come si evince dalla fig. 15.4, le relazioni fra *Opera* e *Thesaurus* sono due.

1. La relazione $X1\text{\u00e9}ClassificatoInX2$ ha dominio *Opera* e codominio *Classe*. La relazione inversa \u00e9 $X2\text{classifica}X1$.
2. La relazione $X1\text{haTermineControllato}X2$ ha dominio *Opera* e codominio *Concetto*. La relazione inversa \u00e9 $X2\text{\u00e9}TermineControllatoDiX1$.

Infine, fra *Opera* e *Fontes* c'\u00e9 una sola relazione.

1. La relazione $X1\text{analizza}X2$ ha dominio *Opera* e codominio *LuogoFonte*. La relazione inversa \u00e9 $X2\text{\u00e9}AnalizzatoDaX1$.

15.2.5. Riepilogo delle relazioni dell'ontologia

Si riportano nella Tab. 1 tutte le relazioni dell'*Ontologia BIA-Net*. Per le relazioni a pi\u00f9 di due posti, l'ultimo dominio \u00e9 indicato nella colonna *CODOMINIO* e gli altri domini sono indicati uno sotto l'altro nella colonna *DOMINIO* secondo l'ordine dei posti. L'ultima riga, in grigio, della tabella indica la relazione `rdfs:subClassOf`, cio\u00e8 l'inclusione fra classi dell'ontologia, che \u00e9 una relazione primitiva di OWL.

Abbiamo sottolineato sopra che la rappresentazione ontologica del *Thesaurus* di *BIA* richiede l'utilizzo di due relazioni ternarie e una quaternaria. Come noto, l'introduzione di relazioni con ariet\u00e0 maggiore di due in linguaggi quali RDF, RDFS (W3C 2004b) e OWL (W3C 2004a) non \u00e9 possibile in modo diretto, in quanto tali linguaggi ammettono soltanto relazioni binarie (Gomez-Perez e Corcho 2002). Questa limitazione \u00e9 usualmente superata adottando particolari pattern ontologici (Gangemi e Presutti, 2009), che permettono di rappresentare una relazione n-aria ($n > 2$) mediante opportune relazioni binarie (W3C 2006). Il pattern ontologico da noi utilizzato per la rappresentazione in OWL delle relazioni di ariet\u00e0 maggiore di due del *Thesaurus* di *BIA* si basa su un nuovo metodo sviluppato dal nostro gruppo di ricerca, che \u00e9 particolarmente adeguato per tener conto dei rapporti di rilevanza semantica tra le diverse entit\u00e0 presenti nell'ontologia (Giunti 2010). Tale metodo, e il relativo pattern ontologico, sono oggetto di un lavoro specifico in corso di ultimazione.

<i>DOMINIO</i>	<i>CO DOMINIO</i>	<i>NOME DELLA RELAZIONE</i>	<i>TIPO DI RELAZIONE</i>	<i>COLORE</i>
Vedetta	Classe	X1èVedettaDiX2	Relazione a due posti	nero
Vedetta	termineV	haTermineV	Funzione totale biiettiva	rosso
Classe	codiceC	haCodiceC	Funzione totale biiettiva	rosso
Classe	Classe	X1èSottoclasseDiX2	Relazione transitiva	marrone
Concetto	Classe	X1èDescrittoreDiX2	Relazione a due posti	nero
Concetto, Concetto	Classe	X1èSussuntoInX2èDescrittore DiX3	Relazione a tre posti	viola
Concetto, Concetto	Classe	X1equivalenteX2èDescrittore DiX3	Relazione a tre posti	viola
Concetto, Concetto, Concetto	Classe	X1equivalenteX2èSussuntoIn X3èDescrittoreDiX4	Relazione a quattro posti	viola
Concetto	termineN	haTermineN	Funzione totale biiettiva	rosso
Opera	Classe	X1èClassificatoInX2	Relazione a due posti	nero
Opera	Concetto	X1haTermineControllatoX2	Relazione a due posti	nero
Opera	titolo	haTitolo	Funzione totale biiettiva	rosso
Opera	Autore	X1haAutoreX2	Relazione a due posti	nero
Autore	nomeA	haNomeA	Funzione totale biiettiva	rosso
Opera	Pubblicazione	X1haPubblicazioneX2	Funzione possibilmente parziale (functional)	verde
Pubblicazione	nomeP	haNomeP	Funzione totale biiettiva	rosso
Opera	numeropagina	haPagina	Funzione possibilmente parziale (functional)	verde
Opera	Anno	X1haAnnoX2	Funzione possibilmente parziale (functional)	verde
Anno	anno_aCdC	haAnno_aCdC	Funzione totale biiettiva	rosso
Opera	testoO	haTestoO	Funzione possibilmente parziale (functional)	verde
Opera	LuogoFonte	X1analizzaX2	Relazione a due posti	nero

Fonte	codiceF	haCodiceF	Funzione totale biettiva	rosso
LuogoFonte	Fonte	X1èLuogoFonteDiX2	Funzione possibilmente parziale (functional)	verde
LuogoFonte	codiceLF	haCodiceLF	Funzione totale biettiva	rosso
LuogoFonte	inscriptio	haInscriptio	Funzione possibilmente parziale (functional)	verde
Testo	subscriptio	haSubscriptio	Funzione possibilmente parziale (functional)	verde
Testo	testoT	haTestoT	Funzione totale biettiva	rosso
Rubrica	rubrica	haRubrica	Funzione totale biettiva	rosso
Tutte le classi dell'ontologia	Tutte le classi dell'ontologia	rdfs:subClassOf	Relazione a due posti primitiva di OWL	blu

Tab. 15.1. Classificazione delle relazioni.

15.3. Conclusioni e linee di sviluppo

Il progetto *BIA-Net: strumenti semantici e applicazioni multimediali per la ricerca nella Bibliotheca Iuris Antiqui* è attualmente in corso, con conclusione prevista nell'ottobre 2013. Come precedentemente detto, avendo già definito lo schema di ontologia qui presentato, il passo successivo consisterà nel suo popolamento, ovvero nell'introduzione delle istanze delle classi e nell'asserzione dei relativi fatti (triple RDF). In questo modo, la base di conoscenze dei diritti dell'antichità attualmente contenuta nei tre archivi di *BIA* sarà infine trasformata in un'unica rete semantica espressa nella forma di un grafo RDF.

Inoltre, in previsione della conclusione del progetto, l'unità di Cagliari, in stretta collaborazione con le altre unità, è al lavoro per la definizione degli scenari futuri, nella prospettiva di un'evoluzione di *BIA* in chiave semantica, in particolare per lo sviluppo di un motore di ricerca e di un browser tramite cui rendere tale base di conoscenze semanticamente interrogabile e navigabile.

Con riferimento al motore di ricerca, le chiavi per l'interrogazione di *BIA* potranno essere di due tipi: un oggetto appartenente a una delle undici classi dell'ontologia (Classe, Vedetta, Concetto, Opera, Autore,

Pubblicazione, Anno, Fonte, LuogoFonte, Testo, Rubrica) o uno dei fatti asseriti esplicitamente dall'ontologia. Il problema consiste pertanto nello sviluppare un metodo che permetta di reperire tutta e sola l'informazione più rilevante relativa a una qualunque chiave di ricerca di questi due tipi. Tale problema è in via di soluzione sulla base di un modello per la misura della rilevanza definito in Giunti (2010). Questa soluzione verrà esposta in successive pubblicazioni. L'implementazione di questo modello potrà portare a un motore di ricerca semantico estremamente raffinato, sensibile al contesto, e capace di filtrare i risultati nei modi più vari e adeguati ai possibili interessi dell'utente.

Con riferimento al browser, esso dovrebbe consentire una navigazione visuale e ipertestuale del grafo dell'ontologia stessa, scoprendo via via l'informazione connessa al punto di partenza della navigazione. Come in un browser standard la navigazione comincia con l'indirizzo di una pagina Web (il suo URL), così nel browser semantico di *BIA* la navigazione comincerà attraverso l'identificazione di un'informazione iniziale. L'informazione iniziale ha una funzione analoga alla chiave di ricerca del motore semantico e, esattamente come quella, può essere di due tipi: un oggetto o un fatto. Una volta che l'utente abbia scelto l'informazione iniziale, il browser si posizionerà su tale informazione, nel senso che mostrerà la porzione di grafo costituita dall'intorno completo dell'informazione iniziale, da cui sarà possibile estendere la navigazione tramite l'espansione del nodo di volta in volta cliccato. Inoltre, sia nel motore di ricerca che nel browser, sarà possibile filtrare le informazioni da visualizzare sulla base delle categorie (classi e relazioni) fornite dall'ontologia.

15.4. Bibliografia

- ANTONIOU G., VAN HARMELEN F. (2008). *A Semantic Web primer*, 2° ed., MIT Press.
- BERNERS LEE T., HENDLER J., LASSILA O. (2001). The Semantic Web. «Scientific American», 2001, May 17.
- GANGEMI A., PRESUTTI V. (2009). *Ontology Design Patterns*. In S. Staab, R. Studer, a c. di, *Handbook of Ontologies*, 2° ed., Springer, pp. 221-243.
- GIUNTI M. (2010). *Grafi pesati e relazioni n-arie: un approccio generale all'organizzazione automatica di dati secondo rapporti di rilevanza*. In P. Storari, E. Gola, a c. di, *Forme e formalizzazioni*, CUEC Editrice, Cagliari, pp. 229-245.
- GOMEZ-PEREZ A., CORCHO O. (2002). *Ontology languages for the Semantic Web*. «IEEE Intelligent Systems», vol. 17, no. 1, 54-60.
- PALAZZOLO N. (1995). *Un Thesaurus per la ricerca sui diritti dell' Antichità. Esperienze e problemi*. «Informatica e diritto», 1995, no. 2, 267-278.
- PALAZZOLO N. (2002). *Biblioteca Iuris Antiqui. Sistema informativo integrato sui diritti dell' Antichità*, include Cd-ROM, Editrice Torre, Catania.
- W3C (2004a). *OWL Web Ontology Language Overview*. W3C Recommendation 10 February 2004. URL: <http://www.w3.org/TR/owl-features/> [ultima visita: 7.01.2013].
- W3C (2004b). *RDF Primer*. W3C Recommendation 10 February 2004. URL: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/> [ultima visita: 8.01.2013].
- W3C (2006). *Defining N-ary Relations on the Semantic Web*. W3C Working Group Note 12 April 2006. URL: <http://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/> [ultima visita: 12.02.2013].

COMITATO EDITORIALE
SAPIENZA UNIVERSITÀ EDITRICE

Coordinatore

ROBERTO NICOLAI

Membri

MAURIZIO DEL MONTE

GIUSEPPE FAMILIARI

VITTORIO LINGIARDI

CAMILLA MIGLIO

DANIELE NARDI

CESARE PINELLI

Delegato del Rettore per l'editoria

LUIGIA CARLUCCI AIELLO

COMITATO SCIENTIFICO
MACROAREA E

Coordinatrice

CAMILLA MIGLIO

Membri

VICENÇ BELTRAN

MASSIMO BIANCHI

ALBIO CESARE CASSIO

EMMA CONDELLO

FRANCO D'INTINO

GIAN LUCA GREGORI

ANTONIO IACOBINI

SABINE KOESTERS

EUGENIO LA ROCCA

ALESSANDRO LUPO

LUIGI MARINELLI

MATILDE MASTRANGELO

ARIANNA PUNZI

EMIDIO SPINELLI

STEFANO VELOTTI

CLAUDIO ZAMBIANCHI

COMITATO SCIENTIFICO
SERIE QUADERNI DIGILAB

Membri

LUIGIA CARLUCCI AIELLO (Sapienza Università di Roma)

TIZIANA CATARCI (Sapienza Università di Roma)

FABIO CIOTTI (Università di Roma Tor Vergata)

GIANFRANCO CRUPI (Sapienza Università di Roma)

DOMENICO FIORMONTE (Università di Roma Tre)

MARIELLA GUERCIO (Sapienza Università di Roma)

GIOVANNI RAGONE (Sapienza Università di Roma)

MARIO RICCIARDI (Università di Torino)

GINO RONCAGLIA (Università della Tuscia)

GIOVANNI SOLIMINE (Sapienza Università di Roma)

Il Comitato editoriale assicura una valutazione trasparente e indipendente delle opere sottoponendole in forma anonima a due valutatori, anch'essi anonimi. Per ulteriori dettagli si rinvia al sito: www.editricesapienza.it

COLLANA CONVEGNI

1. Problemi di campionamento nella ricerca sociale
Enrica Aureli Cutillo
2. L'identità culturale di Roma all'inizio del terzo millennio
Romano Bettini
3. Look Homeward and Forward
A. Lombardo, M. Faraone, M. Melloni, I. Tattoni
4. Living in the city
Eugenio Sonnino
5. Applicazioni di analisi statistica dei dati testuali
Enrica Aureli Cutillo, Sergio Bolasco
6. Raffaello Morghen e la storiografia del Novecento
Ludovico Gatto, Eleonora Plebani
7. Rome and New York City
Comparative Urban Problems at the End of 20th Century
Victor Goldsmith, Eugenio Sonnino
8. L'Italia Restaura
Restauro dei Monumenti e Recupero Urbano in Italia e in Cina
Luigi Gazzola
9. Celestino V. Cultura e società
Ludovico Gatto, Eleonora Plebani
10. Dal diritto di voto alla cittadinanza piena
Marisa Ferrari Occhionero
11. Giornate Europee della Facoltà di Economia
Donatella Strangio
12. Mercato del lavoro e protezione sociale nell'Unione Europea
Giuseppe Burgio, Marina Capparucci, Giuseppe Sancetta ed Enrico Todisco
13. Coesione sociale e sostenibilità nell'Unione Europea
Giuseppe Burgio, Marina Capparucci, Giuseppe Sancetta ed Enrico Todisco
14. La Politica Agricola Comune (PAC) e la gestione dei disastri ambientali
Il ruolo dell'agricoltura
Giuseppe Burgio e Simone Vieri
15. Alla maniera di... Convegno in ricordo di Maria Teresa Lucidi
*Pierfrancesco Fedi, Chiara Silvi Antonini, Paola Mortari Vergara Caffarelli,
Alida Alabiso, Daniela Sadun, Francesco Noci e Tullio Aurizi*
16. Una storia delle scienze per i nuovi saperi. Discussioni e ricerche
Guglielmo Rinziavillo

17. Evolution, Equations and Materials with Memory
Daniele Andreucci, Sandra Carillo, Mauro Fabrizio, Paola Loreti, Daniela Sforza
18. Education and Research without Borders
Benedetta Cassani and Federica Mazzarelli
19. Glimpses of Indian History and Art
Reflections on the Past, Perspectives for the Future
Tiziana Lorenzetti and Fabio Scialpi
20. Giorgio Bazzichelli. L'uomo e lo scienziato
Autori vari
21. Attuazione e sostenibilità del diritto alla salute
Roberto Nania
22. Épicurisme et Scepticisme
Stéphane Marchand & Francesco Verde
23. I musei di chimica e la chimica nei musei della scienza
Atti della giornata studio 23 novembre 2012
Villa Celestina, Pineta Marradi di Castiglioncello (LI)
Luigi Campanella e Valentina Domenici
24. Digital Humanities
Progetti italiani ed esperienze di convergenza multidisciplinare
Fabio Ciotti

Questo volume raccoglie gli atti del primo convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (Firenze 13 e 14 dicembre 2012). L'incontro ha avuto al centro alcuni temi di particolare rilevanza nell'attuale fase di sviluppo delle Digital Humanities: come favorire la convergenza di competenze diverse per sostenere il processo di transizione al digitale dell'intero patrimonio culturale? Come ripensare e riaffermare il ruolo essenziale che l'informatica umanistica è chiamata a svolgere in questo processo di trasformazione culturale? Come valorizzare il contributo culturale e scientifico della ricerca e dei progetti condotti in ambito italiano?

Fabio Ciotti è ricercatore all'Università Roma Tor Vergata dove insegna "Informatica applicata al testo letterario" e "Teoria della Letteratura". Si occupa da anni di diversi temi nel dominio delle Digital Humanities: la codifica informatica dei testi, l'analisi computazionale dei testi, l'applicazione di ontologie e Linked Data agli oggetti culturali. Su questo ha pubblicato numerosi volumi e articoli. È membro del direttivo dell'AIUCD e del Technical Council del TEI Consortium.

ISBN 978-88-98533-27-5



9 788898 533275