

Biblioteche digitali e umanisti informatici: nuove partnership?

Anna Maria Tammaro
(Università degli Studi di Parma, Italia)

Abstract Until now, we have had insufficient collaboration between digital humanists and cultural heritage institutions for the digitisation of resources for research. In the digital library context, scholars have now had new opportunities to create digital resources tailored to their research needs. The vision of the digital library service is no longer related to an editorial model of digital objects, but rather to a technically possible semantic access model. The Round Table has focused on collaboration on Linked Open Data (LOD), which benefits both scholars and professionals.

Sommario 1 Perché gli umanisti informatici hanno bisogno di una Biblioteca digitale? – 2 Perché le biblioteche digitali hanno bisogno degli umanisti informatici? – 3 Conclusioni.

Keywords Digital Library. Linked Open Data. Scholarly primitives. Ontology. Semantic Web. Open Access. Re-use.

Finora è stato molto insolito che i bibliotecari parlassero di digitalizzazione ai Convegni degli umanisti informatici, come anche finora non c'è stata molta collaborazione tra bibliotecari, archivisti, curatori delle istituzioni culturali e docenti delle facoltà umanistiche per la digitalizzazione del patrimonio culturale. È opinione diffusa che i professionisti, come tecnici della digitalizzazione, possano decidere con quali tecniche e criteri attuare la digitalizzazione meglio anche degli utenti. Si pensa infatti che la digitalizzazione sia di per sé una 'cosa buona' e che, una volta che la collezione digitale sia accessibile, gli utenti sicuramente la useranno, nel modo in cui è stata creata dai professionisti.

La pianificazione della digitalizzazione spesso parte quindi dalla creazione di contenuti, perseguendo un modello editoriale, per poi diffondere le collezioni digitali agli studiosi, istruendoli su come dovrebbero essere usate. Gli studiosi generalmente non hanno nessun controllo dell'edizione scelta e della codifica o struttura della fonte, che vincola le funzionalità disponibili, che quindi possono non essere quelle che servono alla ricerca specialistica. Non si cerca di capire prima come gli umanisti informatici vorranno usare la collezione digitale, ma si pensa spesso a un utente generico.

Non c'è poi da meravigliarsi se si scopre che le risorse digitali create dalle biblioteche digitali, anche investendo notevoli quantità di denaro

pubblico, non sono poi usate. Nella letteratura professionale, quando ci si interroga su questo problema, vengono evidenziati possibili ostacoli all'uso delle risorse digitali come la non conoscenza della risorsa digitale, o la mancanza di capacità digitali per usarla o aspetti di difficoltà di accesso come la mancanza di connessione. Potrebbe invece essere che le risorse digitali non sono usate perché non sono utili agli studiosi? Oppure non consentono il tipo di uso che servirebbe?

Nel contesto della biblioteca digitale, gli studiosi sono stati 'liberati' da questi vincoli e hanno avuto nuove opportunità di scelta di accesso alle risorse digitali, fino alla possibilità di creare proprie risorse digitali adeguate ai loro bisogni di ricerca, anche attraverso il ri-uso di collezioni digitali create da biblioteche. La crescita dei contenuti generati dagli studiosi è particolarmente importante per estendere l'accesso a risorse digitali e a metadati scelti in modo appropriato, anche per dare un valore aggiunto a tali risorse attraverso attività come il collegamento, annotazioni, etc. Possiamo quindi evidenziare numerose iniziative di singoli studiosi o di istituzioni di ricerca che hanno costruito collezioni digitali con le funzionalità necessarie per la ricerca, ma queste iniziative - sempre di grande qualità - sono accomunate da una caratteristica che rischia di togliere valore alle risorse stesse: hanno problemi oggettivi di sostenibilità e permanenza in rete per mancanza di risorse e di competenze tecniche.

Può costruirsi un nuovo rapporto di partnership con le biblioteche digitali su queste esigenze e opportunità?

1 Perché gli umanisti informatici hanno bisogno di una Biblioteca digitale?

La biblioteca digitale di antichistica non si limita a dei contenuti digitali a cui si può accedere in linea per una ricerca veloce. Questa è una concezione della biblioteca digitale che deve essere superata: così è da interpretare il significato del titolo dell'intervento di Maurizio Lana «Oltre le biblioteche digitali». La biblioteca digitale non si limita a memorizzare dei contenuti digitali ma rende disponibili funzionalità avanzate. Il valore della biblioteca digitale è tutto in queste funzionalità avanzate, che giustificano l'investimento notevole che richiede la digitalizzazione, facilitando lo studioso a fare quello che altrimenti non sarebbe stato possibile. Come affermava nel 1962 Douglas Engelbart nell'introduzione di *Augmenting Human Intellect* (citato da Maurizio Lana): «Con 'potenziare l'intelletto umano' intendiamo accrescere la capacità di un uomo di affrontare una situazione che presenta un problema complesso, di acquisire comprensione adatta alle sue specifiche necessità, e derivare soluzioni per i problemi». La 'cura digitale' è la funzionalità di base della biblioteca digitale. A differenza delle risorse digitali accessibili dal web, la biblioteca digitale svolge un

insieme di funzioni, tra cui la preservazione delle risorse digitali raccolte, e questa funzione deve essere apprezzata come la più importante, perché al momento solo la biblioteca digitale assicura la permanenza nel tempo dell'accesso alle risorse. Le altre funzioni che sono comprese nella cura digitale sono l'organizzazione dei contenuti digitali, basata sulla selezione dei contenuti, la struttura del testo e l'immissione dei metadati. La cura digitale richiede un assetto tecnologico, di cui la biblioteca digitale deve dotarsi, che mantiene l'accessibilità alle risorse, malgrado l'obsolescenza del software e la necessità di migrazione dei formati. Le biblioteche digitali sono quindi un assetto digitale molto importante per l'infrastruttura di ricerca (io direi unico) e rappresentano uno spazio virtuale connesso a Internet che mantiene la risorsa accessibile nel tempo per l'accesso, l'uso ed il ri-uso. Ciascuno di questi tre concetti: accesso, uso e ri-uso, devono essere ben compresi e hanno delle problematicità che credo tutti gli studiosi hanno già sperimentato.

Cominciamo con l'accesso. I servizi della biblioteca digitale, oltre a identificare e recuperare la risorsa digitale come in tutti i sistemi di Information Retrieval, offrono una serie di funzionalità avanzate dell'interfaccia. Unsworth (2000) ha chiamato queste funzionalità le «scholarly primitives» o priorità di ricerca che devono essere rese disponibili se si vuole essere di supporto allo studioso: 'Annotating', 'Comparing', 'Referring', 'Selecting', 'Linking', 'Discovery'.

Raul Mordenti nella sua presentazione «Alcuni problemi teorici e pratici posti dalle *digital libraries*» ha evidenziato le difficoltà che ancora si frappongono a un servizio di biblioteca digitale per l'antichistica. Il metodo di ricerca predominante degli umanisti informatici è complesso. Lo studioso ha bisogno di numerose fonti e opere da consultare insieme, e non di eseguire ricerche per parole chiave, mentre la maggior parte invece dei sistemi informativi e delle risorse digitali prevedono questo tipo di ricerca. Lo studioso di antichistica deve seguire le note nel testo che sta leggendo o cercare informazioni collegate in altri testi connessi al primo. Gli umanisti informatici hanno cioè bisogno di fare ciò che Bates chiama 'raccolta di bacche' - in altre parole, lo studioso vuole estrarre citazioni da testi che sono particolarmente attinenti all'argomento di studio, piuttosto che citare tutto quello che è stato scritto sull'argomento (Bates 1989). Hanno anche bisogno di una più ampia gamma di testi, in termini di data di pubblicazione e tipologia di pubblicazioni: invece di leggere articoli di periodici degli ultimi cinque anni, gli umanisti informatici hanno bisogno di consultare libri stampati o manoscritti che hanno centinaia di anni, così come immagini, film, musica, mappe, oggetti museali e altre tipologie di materiale di origine storica. Spesso devono rileggere o riesaminare fonti in un modo complesso, piuttosto che fare la ricerca di documenti digitali per informazioni fattuali.

Un modello di biblioteca digitale basata sulla collaborazione è *digilibLT*, Biblioteca digitale sul latino tardo, presentata da Maurizio Lana nella

Tavola rotonda. La biblioteca digitale *digilibLT* è nata proprio per il ri-uso e distribuisce liberamente sotto licenza Creative Commons le opere del latino tardo in vari formati tra cui ePUB, il formato aperto degli ebook. Il gruppo di lavoro è rappresentato da latinisti insieme a bibliotecari ed informatici.

In conclusione, cercando di rispondere alla domanda: perché gli umanisti informatici hanno bisogno della biblioteca digitale? Sulla base dei progetti sinteticamente descritti dai partecipanti alla Tavola rotonda la prima risposta importante credo che sia 'per la sostenibilità e l'accesso permanente delle risorse digitali'. Un altro motivo è che la biblioteca digitale crea uno spazio virtuale per l'uso, il ri-uso e stimola la collaborazione interdisciplinare. Questa collaborazione ancora non è diffusa, ma viene stimolata dall'Unione Europea e presto sarà comunque evidente dai risultati dei progetti europei e nazionali che sono stati descritti.

C'è un ostacolo però: manca l'accesso aperto alle risorse digitali. I partecipanti alla Tavola rotonda sono stati tutti d'accordo nel definire l'accesso aperto - definito nel modo più ampio - come 'pre-requisito' per la biblioteca digitale d'antichistica.

2 Perché le biblioteche digitali hanno bisogno degli umanisti informatici?

Abbiamo evidenziato come il contesto di riferimento delle biblioteche digitali è diverso da quello delle biblioteche tradizionali e la differenza sta nel ruolo attivo che hanno gli studiosi. Gli umanisti informatici sono creatori essi stessi di contenuti digitali e hanno la possibilità di collaborare alla costruzione e gestione della biblioteca digitale. La prima domanda che era stata fatta ai partecipanti alla Tavola rotonda era: quali sono i contenuti digitali di cui voi siete i creatori e quanto siano *open*, cioè aperti, alla condivisione e collaborazione con altri studiosi anche di discipline diverse?

Un campo di applicazione dove l'apertura sembra avere meno ostacoli è quello dei metadati e delle ontologie. Usando la definizione di Maurizio Lana (citata da Maurizio Lana nel suo intervento durante la Tavola rotonda)

un'ontologia è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di conoscenza, organizzata in una gerarchia di categorie; tale strutturazione permette di costruire ed effettuare per mezzo del computer semplici ragionamenti di tipo formale/deduttivo.

Il Progetto Geolat (Geography for Latin Literature) è un esempio di ontologia che permette a studiosi e cittadini alla ricerca dei nomi geografici contenuti nei testi letterari di scoprire nuovi modi per valutare e leggere i rapporti che legano i testi geografici.

Nel caso dei dati aperti e delle ontologie, la collaborazione tra umanisti informatici e biblioteche digitali diviene necessaria (non solo desiderabile). L'uso dei linguaggi aperti del web è quindi la conseguenza necessaria del bisogno di collaborare.

Annotare i testi utilizzando l'ontologia è un'operazione complessa, che può essere fatta solo da studiosi – avverte Maurizio Lana – perché occorrono elevate conoscenze della lingua e della storia. Ad esempio in Geolat occorre:

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

Anche Anna Lucarelli è d'accordo sul potenziamento della collaborazione tra bibliotecari e studiosi, partendo dal punto fondamentale che le biblioteche digitali si stanno trasformando da *repositories* di testi a *semantic digital libraries*. Partendo dall'esperienza della trasformazione del Soggettario della Biblioteca Nazionale Centrale di Firenze in relazione alle prospettive introdotte dal web semantico e dai *linked data*, è aumentata la sua portata multidisciplinare. La Biblioteca Nazionale ha la responsabilità istituzionale di mantenere uno strumento di indicizzazione semantico, e il passaggio più opportuno sembra ora quello di avviare una partnership tra il Nuovo Soggettario e gli strumenti di indicizzazione semantica specialistici che sono stati realizzati dagli umanisti informatici, nell'ottica del web semantico.

In conclusione, il progetto cooperativo per una Biblioteca digitale di antichistica oggi non può che passare per i Linked Open Data (LOD) che permettono di collaborare a livello internazionale e interdisciplinare, anche a distanza nel tempo e nello spazio, lasciando a ciascuno dei partner libertà di impianto dei contenuti ma adottando una modalità di interoperabilità molto forte. Per ottenere questo risultato non è sufficiente basarsi sulla tecnologia del web semantico, ma occorre costruire le ontologie su uno schema logico, come ad esempio quello presentato da Marco Giunti per la Biblioteca Iuris Antiqui in questo Convegno. Vi cito questo caso, perché credo che il problema degli Open Data dovrebbe essere al centro della discussione per la costruzione di una Biblioteca digitale di antichistica. Questa discussione dovrà necessariamente avere un focus su uno schema logico, anche se si potranno avere esperienze diverse.

3 Conclusioni

Possiamo concludere dopo la Tavola rotonda del Convegno di Venezia che le funzionalità disponibili nelle attuali biblioteche digitali non sono quelle descritte come prioritarie nell'elenco di Unsworth.

Come potrebbe migliorare la ricerca negli studi di antichistica? La visione del servizio di accesso della biblioteca digitale non è più legata a un modello editoriale di rappresentazione digitale di oggetti in rete, ma piuttosto a un modello di accesso semantico tecnicamente possibile. La base per una biblioteca digitale per l'antichistica sarà costruita dalla forte motivazione a collaborare e a condividere le risorse digitali costruite finora da istituzioni culturali e umanisti informatici, contribuendo in questo modo a una migliore produttività della ricerca.

Sui bisogni di un progetto cooperativo abbiamo ascoltato le relazioni del Convegno e della Tavola rotonda, per concludere che il primo progetto comune di collaborazione dovrebbe essere relativo al ruolo di Linked Open Data (LOD) nello studio del mondo antico. Il termine, LOD comprende approcci alla pubblicazione di risorse digitali che enfatizzano la stabilità, l'accesso granulare al contenuto intellettuale tramite pubblico URI, e la riutilizzabilità come definito mediante licenze che permettono ulteriori copia di materiali disponibili. In una precedente Tavola rotonda sui Linked Data durante l'ultimo Convegno AIUCD, avevo definito i due approcci che si confrontano oggi nel contesto della Biblioteca digitale 'Paradigma chiuso' e 'Paradigma aperto'. I due paradigmi definiscono non tanto la struttura dei dati ma l'apertura alla condivisione e al ri-uso di questi al fine di facilitare la ricerca collaborativa e quindi migliori risultati.

Le biblioteche digitali possono avere un ruolo molto importante nel facilitare la ricerca degli umanisti informatici. Il loro ruolo sta nel promuovere l'accessibilità e la conservazione dei contenuti digitali, attraverso la diffusione dei contenuti creati dagli umanisti informatici, fino alla collaborazione per Linked Open Data e altri strumenti di funzionalità avanzate, come annotare, comparare, connettere, ecc. Mi chiedo perché questa collaborazione ancora non ci sia per le discipline antichistiche, ma sicuramente le tecnologie ci sono, e le metodologie di ricerca e gli strumenti potrebbero essere in grado di adattarsi ai cambiamenti nello scenario in continua evoluzione delle università e della ricerca universitaria.

Ancora oggi, dopo tre anni dalla fine del Convegno, la comunità dei curatori delle istituzioni culturali esperta di Linked Open Data non collabora con la comunità degli umanisti informatici, come ci si proponeva nella conclusione della Tavola rotonda, incoraggiando l'interconnessione e l'interoperabilità delle risorse digitali aperte.

Finora l'adozione di Linked Open Data da parte della comunità degli umanisti informatici è stata infatti modesta. Ancora possiamo parlare solo di 'promesse' di come questa tecnologia potrebbe essere un mezzo per la

pubblicazione e lo scambio di dati tra i progetti degli umanisti informatici e quelli delle istituzioni culturali. Siamo tuttavia ancora convinti che la creazione di ontologie potrebbe diventare un mezzo per condividere concetti astratti complessi e collaborare per approfondire lo studio dei temi di ricerca degli umanisti informatici.

Alcuni degli ostacoli che possiamo elencare circa questa possibile collaborazione tra biblioteche digitali e gli umanisti informatici sono l'insufficienza di strumenti collaborativi a disposizione delle biblioteche digitali, la mancanza di standardizzazione all'interno dei domini di studio degli umanisti informatici, la preoccupazione per l'applicabilità di ontologie specialistiche alle esigenze di un'utenza generica e soprattutto la disponibilità di fonti di dati aperte e 'pulite' cioè rispondenti alle esigenze degli umanisti informatici e con cui iniziare la collaborazione.

Il Convegno di cui si pubblicano gli atti ha avuto il merito di essere il primo sul tema e ha sicuramente aumentato la consapevolezza dei nuovi sviluppi possibili (sia per le risorse digitali che per gli strumenti collaborativi) con la condivisione delle migliori pratiche.

Bibliografia

- Bates, Marcia J. (1989). «The Design of Browsing and Berrypicking Techniques for the Online Search Interface» [online]. URL <https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html> (2017-10-25).
- CENDARI, Collaborative European Digital Archive Infrastructure. URL <http://www.cendari.eu> (2017-10-25).
- COST Action IS1005. «Medieval Europe - Medieval Cultures and Technological Resources» [online]. URL http://www.cost.eu/COST_Actions/isch/Actions/IS1005 (2017-10-25).
- DARIAH, Digital Research Infrastructure for the Arts and Humanities. <http://it.dariah.eu/sito/> (2017-10-25).
- Engelbart, Douglas C. (1962). «Augmenting Human Intellect. A Conceptual Framework» [online]. URL <http://www.dougelbart.org/pubs/augment-3906.html> (2017-10-25).
- TRAME, Text and Manuscript Transmission of the Middle Age in Europe. URL <http://trame.fefonlus.it/trame/index.html> (2017-10-25).
- Unsworth, John (2000). «Scholarly Primitives. What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This» [online]. URL <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html> (2017-10-25).

