

Digitalizzazione e Indicizzazione delle Storie Aquilane

Alessandro Celi, Romina Eramo, and Marco Segala

Università degli Studi dell'Aquila,
via Vetoio, 67100, L'Aquila, Italy
{nome.cognome}@univaq.it

Sommario Il mondo delle biblioteche e dei servizi bibliotecari e la nuova realtà rappresentata dall'ecosistema informativo digitale hanno oggi più che mai bisogno di una stretta collaborazione. Questo articolo presenta un progetto di digitalizzazione libraria che ha come scopo la circolazione della cultura, dedicando particolare attenzione al ruolo delle biblioteche e alla storia dei testi in essa contenuti. Il progetto è dedicato alla digitalizzazione della Biblioteca Salvatore Tommasi della città dell'Aquila e, in particolare, dei volumi manoscritti degli annali di Anton Ludovico Antinori (sec.XVIII). Il progetto offre uno spunto di riflessione in ambito accademico sulla natura e sul possibile ruolo della digitalizzazione nella trasmissione della cultura, ma anche relativamente alle conseguenze della rivoluzione digitale nel mondo bibliotecario.

Keywords: Digital Humanities, Digitalizzazione libraria, Text Encoding Initiative

1 Introduzione

La digitalizzazione sta modificando profondamente il ruolo delle biblioteche. Le nuove tecnologie offrono possibilità mai esplorate, in grado di modificare radicalmente la vita degli individui e le relazioni sociali e con l'ambiente; allo stesso tempo, però, esse sfidano nozioni comuni e intuizioni morali condivise, e per questo suscitano interrogativi che meritano di essere indagati. Alla base del progetto descritto in questo articolo c'è l'idea che questioni, risposte e attività che vertono sull'uso innovativo della tecnologia necessitano di essere divulgate perché alimentino la cultura condivisa e la consapevolezza dei cittadini.

La Biblioteca Salvatore Tommasi della città dell'Aquila possiede importanti collezioni manoscritte e librerie che da sempre hanno costituito motivo per la sua presenza come protagonista nel panorama culturale del territorio. Il sisma del 2009 e la sistemazione della Biblioteca in una zona periferica hanno drasticamente limitato la sua presenza nella vita culturale della città e hanno reso meno fruibile il suo patrimonio. Il progetto Digitalizzazione e Indicizzazione delle Storie Aquilane (DigISA), grazie al sostegno di Fondazione TIM e Regione Abruzzo, mira sia a restituire centralità e peso alla Biblioteca nel contesto culturale locale,

sia a rendere fruibile un prezioso patrimonio librario altresì difficilmente accessibile. In particolare, si sceglie di digitalizzare una parte del patrimonio librario della Biblioteca attualmente concessa alla consultazione con limiti e prescrizioni per salvaguardarne la conservazione: i volumi manoscritti degli Annali di Anton Ludovico Antinori [1] finora mai pubblicati (Corografia, vol. XXXII-XLII.2). Si tratta di testi di primaria importanza per la comprensione della storia regionale che da decenni sono diventati le fonti principali della ricerca storica.

Il progetto comprende, da un lato, l'uso delle più recenti tecnologie nell'ambito delle digital humanities (DH) [3] per la valorizzazione di fonti fondamentali per la comprensione della storia territoriale, così da offrire ai cittadini la possibilità di riappropriarsi delle narrazioni che hanno definito la loro identità comune e la loro cultura di appartenenza; e dall'altro l'apertura di un cantiere di riflessione sulle sfide concettuali, morali e valoriali che lo sviluppo tecnologico presente pone agli individui nel contesto sociale, ambientale e culturale. Questa duplice declinazione della tecnologia si utilizza per determinati scopi culturali, ma diventa essa stessa oggetto di riflessione per una ridefinizione dei contenuti stessi di ciò che chiamiamo cultura.

Il progetto si avvale delle competenze presenti nell'Università dell'Aquila negli ambiti dell'informatica, della ricerca storica, filosofica e delle DH per avviare un processo di arricchimento e diffusione della cultura. Oltre alla produzione di materiale fruibile per iniziative di ricerca e diffusione del patrimonio culturale, il progetto prevede la formazione di laureandi e giovani studiosi alle metodologie e alle pratiche di conservazione, arricchimento e diffusione delle risorse culturali attraverso strumentazione tecnologica, software dedicato e applicazioni delle DH. Il caso della Biblioteca Tommasi ambisce a rappresentare un esempio di digitalizzazione libraria finalizzato alla diffusione del sapere prodotto, attraverso una piattaforma web dedicata che mira ad avere un importante impatto mediatico.

2 Tecnologia e Cultura

Uno dei principali interrogativi nella società di fronte all'innovazione tecnologica e al dilagare di oggetti e processi tecnologici nella quotidianità è dove tutto questo ci porterà. La tecnologia è, per la quasi totalità della popolazione, produttrice di strumenti e attività che restano misteriosi – scatole nere dove nessuno può guardare perché non saprebbe proprio cosa guardare. Non capiamo quasi nulla di quello che, tuttavia, è diventato familiare; il progresso tecnologico apre possibilità quasi mai richieste e spesso impensate; stringiamo nelle mani oggetti facili da usare che però ci ricordano ogni giorno la nostra incapacità a comprendere il nostro stesso presente e, talvolta, a diffidare del futuro.

Pensare la tecnologia significa interrogarsi sul senso e sul valore di quello che rappresenta e ci offre. Oggi si legge frequentemente di intelligenza artificiale che interagisce sempre più autonomamente con l'ambiente e gli esseri umani – dai robot alle automobili a guida autonoma. Ci sono molte descrizioni, qualche richiamo alla prudenza, ma non c'è una riflessione sistematica su cosa significhi programmare un sistema artificiale a interagire con il mondo e i suoi abitanti. Pur

senza una coscienza, sarà coscienzioso? Gli si potranno insegnare valori morali, oltre che leggi, codici e regole?

D'altra parte, una riflessione teorica sulla tecnologia appare molto più motivata e fondata se sviluppata in un ambiente dove la tecnologia si sposa con la tradizione e contribuisce al suo rilancio e alla sua presenza nella società. È rischioso parlare di tecnologia senza in qualche misura vederla all'opera anche dall'interno – e non solo come “prodotto” da utilizzare o far funzionare. Per questo motivo si mette in opera un progetto nell'ambito delle DH, che negli ultimi anni ha realizzato una delle più stimolanti intersezioni tra tecnologia e produzione culturale in ambito umanistico.

Avere a disposizione in formato digitale questi documenti permette di valorizzare e rivitalizzare la presenza culturale della biblioteca con il progetto tecnologico innovativo di una piattaforma Internet dove tali documenti non sono soltanto presentati ma anche messi a disposizione per essere “lavorati”. Tale piattaforma è infatti corredata di strumenti per la diffusione e la valorizzazione del patrimonio digitalizzato secondo lo stato dell'arte nell'ambito delle DH. In particolare essa offre l'integrazione delle immagini con il database per la ricerca e il collegamento di metadati (dati utilizzati per descrivere altri dati) ma soprattutto ingloba un software per la trascrizione e l'annotazione dei facsimili da parte dei visitatori del sito e degli storici. In altre parole, la Biblioteca mette a disposizione dei cittadini uno strumento tecnologico raffinato che permette di riscoprire un “giacimento” culturale del territorio e la possibilità di riappropriarsi di un'eredità culturale spesso trascurata e talvolta dimenticata. Attraverso di esso, la Biblioteca organizza e sostiene un'iniziativa collaborativa (basata sul microvolontariato) per la trascrizione e l'annotazione delle fonti della storia aquilana, una volta disponibili online con la digitalizzazione dei fondi manoscritti. In questo modo la digitalizzazione delle cronache aquilane diviene strumento di attività culturale che coinvolge i cittadini e che scorre in due sensi: una fruizione del patrimonio culturale che a sua volta arricchisce il patrimonio stesso.

Si tratta di una forma di valorizzazione del patrimonio culturale già sperimentato in altri paesi europei (Francia, Gran Bretagna, Irlanda, ad esempio). L'approdo su web delle pagine dedicate alla storia dell'Aquila diventerà occasione per riavvicinare i cittadini alla loro storia. Si offrirà ai “volontari” la possibilità di trascrivere, secondo gli standard codificati nell'ambito delle DH, le pagine delle cronache aquilane rese disponibili dal progetto di digitalizzazione. Si tratta non solo di promuovere la fruizione della cultura ma anche di chiamare i cittadini a produrla.

3 Obiettivi del Progetto DigISA

Lo scopo principale del progetto è quello di rendere accessibile, attraverso la conversione in un formato digitale, un “nucleo centrale” del patrimonio librario della Biblioteca “Salvatore Tommasi” dell'Aquila, attualmente concesso alla consultazione con limiti e prescrizioni per salvaguardarne la conservazione. La digitalizzazione mira a rendere detto patrimonio fruibile anche al più remoto

degli utenti, favorisce la più ampia conoscenza in ambito nazionale e internazionale dello stesso come pure di tutte quelle manifestazioni del pensiero e dell'arte veicolate dal libro, quale cassa di risonanza della cultura in tutte le sue forme espressive. In tal modo testi antichi e preziosi manoscritti saranno comunque preservati da manipolazioni e conservati a futura memoria.

Del materiale selezionato si delocalizza il contenuto, creando surrogati (immagini) accessibili agli utenti tramite computer e come tali facilmente gestibili e spostabili tra server e computer. Dunque valorizzazione, sicurezza e garanzia conservativa, ma anche fruibilità del variegato patrimonio anche dal più remoto degli utenti, soddisfacendo le aspettative dell'utenza alla più ampia conoscenza delle manifestazioni del pensiero e dell'arte veicolate dal libro. La realizzazione del progetto, oltre a consentire l'acquisizione e la conservazione di materiale digitale, permette all'utenza la ricerca, la visualizzazione, il salvataggio e la stampa, garantendone l'autenticità.

3.1 Selezione del materiale da digitalizzare

La selezione del materiale è strettamente correlata con le finalità che la digitalizzazione intende perseguire. Va considerato, oltre la fattibilità tecnica e il valore aggiunto che ne deriva, anche l'entità dei costi/benefici e la loro sostenibilità nel tempo. Possibili criteri di selezione sono:

- valore storico e culturale;
- unicità e rarità;
- accesso limitato e con restrizioni per salvaguardarne lo stato di conservazione;
- alto interesse di studiosi;
- volontà di favorire l'accesso in linea a collezioni virtuali, in modo da aumentare l'interesse per materiali poco o affatto conosciuti;
- mancanza di vincoli legali o permessi di digitalizzazione ottenuti;
- specificità culturali, geografiche, storiche, economiche del territorio sul quale incide l'opera.

3.2 Piattaforma web

La piattaforma del progetto DigISA è sviluppata come un'applicazione web realizzata con lo scopo di gestire il processo di digitalizzazione dei manoscritti dalla scansione dell'originale, la successiva trascrizione fino alla pubblicazione in rete. Si divide nelle seguenti componenti:

Gestore della digitalizzazione: Il manoscritto è acquisito dal sistema sotto forma di immagini digitali ad alta risoluzione attraverso scanner planetari. Ogni manoscritto è formato da più acquisizioni (ogni immagine rappresenta una singola pagina). La digitalizzazione viene controllata da supervisor all'acquisizione per assicurarne la correttezza (ad esempio, in accordo con standard richiesti) e la qualità. L'immagine acquisita viene memorizzata all'interno del sistema ed assegnata all'opera di riferimento, corredandola di opportuni metadati.

Gestore delle trascrizioni: Il manoscritto così acquisito deve essere trasformato in un testo digitale; ciò avviene attraverso operazioni di trascrizioni in formato TEI (Text Encoding Initiative) [7]. Lo schema di codifica della TEI, basato sulla sintassi dello Standard Generalized Markup Language (SGML, ISO 8879)[8], è rivolto a chi intende produrre e diffondere testi in formato elettronico, in particolare nel dominio umanistico. Esso consente, attraverso appositi *tag*, di rappresentare, la struttura astratta di varie tipologie testuali (testo in prosa, testo poetico, testo drammaturgico, fonte manoscritta, etc.), oltre alle caratteristiche intrinseche del contenuto dell'opera. Tali trascrizioni sono digitate manualmente (la natura del testo rende inutilizzabili strumenti di acquisizione automatica) attraverso un text editor TEI integrato nell'applicazione web messa a disposizione dei trascrittori.

Gestore della pubblicazione: I manoscritti, una volta digitalizzati e superata la fase di revisione delle immagini, vengono pubblicati sul sistema e resi accessibili agli utenti della piattaforma. Le corrispondenti trascrizioni sono pubblicate successivamente, dopo la validazione della stessa. L'intero processo di trascrizione è supervisionato da esperti della materia che correggono ed approvano le trascrizioni pervenute, il tutto sempre attraverso la piattaforma web.

3.3 Formazione

I modernizzati servizi bibliotecari riflettono la transizione del modo in cui i servizi sono erogati agli utenti del mondo bibliotecario. Il progetto DigISA è anche un'importante occasione di formazione per giovani e studenti universitari. All'interno del progetto sono state finanziate, anche grazie al contributo della Fondazione TIM, un numero di borse di studio rivolte agli studenti dell'Università dell'Aquila. In questo modo gli studenti diventano parte attiva dell'intero processo, dalla scansione, all'indicizzazione, alla trascrizione, fino alla pubblicazione e consultazione. Il modello di apprendimento *learnig by doing* permette l'aumento delle conoscenze globali e l'acquisizione di un know-how riutilizzabile in futuro. Di seguito sono elencati alcuni degli argomenti con cui gli studenti si confronteranno:

- gestione del patrimonio librario, in particolare quello antico;
- sistemi informativi e processo di digitalizzazione, relativa verifica di qualità;
- sistemi di trascrizione e linguaggi di *Markup Language* [8];
- indicizzazione e archivio elettronico, revisione dell'elaborato digitalizzato al fine di renderlo consultabile in modo efficace ed efficiente, fruizione on-line;
- gestione e conservazione dei prodotti digitali;
- sostenibilità del servizio;
- diritti di proprietà e copyright;
- modelli di business e print on demand.

4 Processo di Digitalizzazione

La strategia vincente di un progetto di digitalizzazione sta nell'accessibilità che passa soprattutto attraverso un uso ottimale di metadati, vera e propria carta d'identità dei documenti, che ne consente il facile reperimento in Internet e nei cataloghi on line. Le caratteristiche di una buona risorsa elettronica si basano sull'ipertestualità, la capacità migratoria del contenuto, la pluriaccessibilità della risorsa. Per ottenere tutto ciò bisogna seguire il più possibile standard d'uso e best practices approvati dalla comunità scientifica internazionale. Catalogazione, scansione, produzione dei metadati, indicizzazione, controllo di qualità, trascrizione, mantenimento nel tempo della collezione (migrazione dei dati per l'evoluzione delle tecnologie) sono parti importanti di un progetto globale. L'acquisizione digitale costituisce la fase centrale e più delicata del progetto, ma la gestione dei prodotti digitali e la conservazione comportano particolari accortezze e adeguati investimenti economici. Nel perseguire risultati concreti, realistici e sostenibili, la digitalizzazione comprende quattro distinte attività: a) scansione ottica dei supporti cartacei, b) Monitoraggio e controllo di qualità, c) indicizzazione metadati, catalogazione e d) riconoscimento e trascrizione testi in caso di manoscritti.

4.1 Scansione ottica dei supporti cartacei

L'esigenza di garantire una buona qualità della digitalizzazione e la salvaguardia degli originali, inducono ad una attenta scelta della metodologia di acquisizione della strumentazione (sistema di ripresa, di illuminazione, software, tempi di acquisizione ed elaborazione delle immagini, occupazione della memoria nei supporti di memorizzazione). La tipologia di Hardware utilizzato in questi contesti si basa sulla tecnologia a scanner planetario. Tale scanner si presta all'uso specialmente per documenti rilegati e documenti più grandi del formato A3, quali libri, atlanti, opere d'arte su carta, materiali grafici, mappe, manoscritti, pergamene, stampe, ecc. Per i volumi antichi e preziosi, e al fine di evitarne il deterioramento, è opportuno l'utilizzo di un sistema di illuminazione a lampade fredde e a bassissima emissione di raggi IR e UV. Devono essere valutate le esigenze di acquisire anche il retro o altri lati del supporto e se le varie facce formino un'immagine unica o più immagini separate; come pure si deve valutare il formato minimo e massimo dei supporti cartacei, nonché le condizioni medie di conservazione con particolare riferimento agli inchiostri. Le immagini sono prodotte in diversi formati la cui risoluzione di scansione accettabile varia tra i 200–400 ppi (pixel per inch). In merito ai formati prescelti si adottano formati standard per consentire la massima interoperabilità dei prodotti digitali. Vengono prodotte immagini destinate alla conservazione (master), memorizzate in formato non compresso, oppure in formato compresso reversibile (senza perdita); per le immagini destinate alla fruizione vengono usati metodi multirisoluzione, che consentono di graduare la qualità dell'immagine visualizzata da livelli assai bassi fino alla massima qualità. Alcune buone regole da seguire sono:

- Ridurre al minimo le manipolazioni degli originali;
- Usare formati standard evitando quelli proprietari che esigono specifici software per la fruizione;
- Creare master (file ad alta risoluzione) da sottoporre a strategie di conservazione, dai quali sarà possibile creare file derivati per la visualizzazione sul web, per evitare pluralità di digitalizzazione nel tempo.

4.2 Monitoraggio e controllo di qualità

Il monitoraggio temporale delle attività è costante per il superamento di eventuali criticità e la formulazione di proposte migliorative in corso d'opera. Il sistema di controllo della qualità delle immagini è finalizzato ad assicurare la buona leggibilità a video di tutto il contenuto informativo presente negli originali cartacei; presuppone la creazione di un prototipo del prodotto sulla base di un campione significativo del catalogo; l'approvazione formale del prototipo costituirà riferimento di qualità per il processo di scansione. Possono essere utilizzati anche sistemi di controllo aggiuntivi, come programmi automatici di individuazione delle immagini difettose basati su un registro quotidiano dell'attività. Il controllo di qualità, effettuato durante l'intero processo di digitalizzazione, prevede verifiche, effettuate su file master, che hanno per effetto i seguenti elementi:

- correttezza dell'inquadratura e della esposizione;
- colore, profondità, dimensione e formato digitale, eventuali riflessi o deformazioni;
- controllo immagine a video e resa dell'immagine attraverso prove di stampa;
- correttezza dei metadati associati all'immagine e del processo di OCR quando attivo.

4.3 Indicizzazione e catalogazione

Le collezioni di documenti digitali possono essere considerate come nuovi "oggetti culturali". Essi possono derivare da originali analogici o essere digitali nativi, e possono avere molteplici tipologie: immagini statiche, documenti in formato testo, suono, documenti audiovisivi. Al fine di una efficace identificazione e gestione, essi richiedono la definizione di campi descrittivi articolati e versatili. Tali metadati sono strutturati secondo specifici standard e rappresentano l'identificativo delle collezioni digitali.

La creazione e l'organizzazione di metadati è da sempre al centro delle attività di una biblioteca, e diverse sono le proposte per la loro tassonomia. Nel nostro caso il ruolo fondamentale è stato riservato ai metadati di tipo *descrittivo* [4] che permettono l'identificazione ed il recupero degli oggetti digitali attraverso delle descrizioni normalizzate dei documenti. Tali informazione permettono tra l'altro l'interoperabilità tra archivi dentro e fuori il sistema informativo della Biblioteca.

Di seguito elencati alcuni metadati che possono essere associati a ciascun file-immagine:

- informazioni tecniche di base: DPI ripresa, formato grafico, dimensioni immagine, giorno e ora scansione;
- informazioni di base sui materiali: inventario, collocazione del materiale;
- informazioni specifiche per agevolare la ricerca delle informazioni: oggetto, titolo, autore ecc.;
- Indicizzazione del testo e dei metadati utilizzando specifici software;
- controllo di qualità e gestione dei contenuti con software di tipo collaborativo, ad esempio CMS [5] ;
- luogo di pubblicazione dei dati, (locale e sul portale web dedicato), ovvero diritti di proprietà e copyright.

Particolarmente delicata è la configurazione del *server*: archivio elettronico in cui memorizzare i materiali una volta digitalizzati. Un server deve essere strutturato in modo che, oltre a conservare i dati in maniera efficiente, permetta che i documenti siano indicizzati con criteri atti a facilitarne l'individuazione e il recupero per testi, per grafica, per aree di ricerca, per contenuti o destinazione. Al fine di ottimizzare le operazioni di indicizzazione e ricerca, questo progetto utilizza due diversi sistemi: per la memorizzazione sono stati utilizzati dei database di tipo relazionali [10], mentre per la fase di indicizzazione e ricerca sono stati utilizzati veri e propri motori di ricerca, nel nostro caso SOLR [9].

4.4 Riconoscimento e Trascrizione

Una volta acquisita l'immagine del documento attraverso il processo di scansione si procede alla digitalizzazione del testo in esso contenuto. Tale procedura può essere resa automatica attraverso operazioni di OCR [6] quando il documento è stato originato da un processo di stampa. Nel caso di manoscritti-come nel nostro caso- tale procedimento è inefficace, e si rende pertanto necessario l'intervento umano. Infatti solo le operazioni di trascrizioni manuale permettono di interpretare i manoscritti nel modo corretto. Questi tipi di trascrizione sono effettuati utilizzando un apposito *Markup Language* chiamato TEI (Text Encoding Initiative)[7] che è ormai diventato uno standard nell'ambito delle DH.

5 Conclusioni e Sviluppi Futuri

Il presente progetto di digitalizzazione si propone di fare da apripista al recupero e alla fruizione di altri fondi importanti conservati alla Biblioteca Tommasi (corali medievali miniati, incunaboli, cinquecentine). L'auspicio è di creare un diffuso interesse presso gli studenti e la cittadinanza di un nuovo modo di fare cultura mentre se ne fruisce. Naturalmente c'è l'auspicio di sviluppare collaborazioni con altre istituzioni e biblioteche che promuovono esperienze similari. Inoltre, potranno altresì prevedersi collaborazioni atte a tutelare la conservazione e promuovere la diffusione di ingenti e preziosi patrimoni librari del territorio, come ad es. quelli religiosi, di cui la provincia è particolarmente ricca. Altro filone di ricerca da intraprendere è quello relativo alle trascrizioni oggi manuali

e collaborative. Attraverso strumenti di intelligenza artificiale si può, una volta prodotto un prototipo di trascrizione necessario per l'apprendimento del sistema, automatizzare l'intero processo di riconoscimento del testo manoscritto.

Riferimenti bibliografici

1. Orazio D'Angelo, I manoscritti dell'Antinori, in Anton Ludovico Antinori e il 2. centenario della sua nascita , pubblicazione della Società di storia patria negli Abruzzi, pp. 105-118, L'Aquila : Tip. di A. Perfilia, 1904.
2. Extensible Markup Language (XML) (1996). W3C Working Draft 14-Nov-96 Editors: Tim Bray (Textuality), C. M. Sperberg-McQueen (University of Illinois at Chicago)
3. Berry, D. M., ed. (2012). Understanding Digital Humanities, Basingstoke: Palgrave Macmillan.
4. Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche, www.iccu.sbn.it/
5. Ulrich Kampffmeyer (2006). ECM Enterprise Content Management, Hamburg, ISBN 3-936534-09-8
6. Schantz, Herbert F. (1982). The history of OCR, optical character recognition. [Manchester Center, Vt.]: Recognition Technologies Users Association. ISBN 9780943072012.
7. Burnard, L., Sperberg-McQueen, C.M. (2016). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Text Encoding Initiative Consortium, Version 3.0.0. revision 89ba24e
8. Smith, Joan M and Stutely, Robert (1988). SGML: the user's guide to ISO 8879. Halsted Press
9. Apache SOLR Project webpage <http://lucene.apache.org/solr/>
10. Codd, E. F. (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. 13 (6): 377-387. doi:10.1145/362384.362685.