

Un progetto di sviluppo di *Digital Philology*: didattica e ricerca

di *Sabrina Galano e Maria Senatore Polisetti*¹

Abstract

The aim of this paper is to propose a new methodological approach to the teaching of literature through the acquisition and use of the latest information technology. We imagine a new key to reading texts that will enable the works, their authors and the figures quoted to emerge and be contextualized in the light of the close network of data exchange and preferential pathways offered by the web. Research and teaching must be reformulated on the basis of the new technologies dedicated to the organization of information on the internet (the creation of digital display cases and databases for the consultation of texts) so as to borrow from similar sectors the ability to transmit and codify data, that is both structured and unambiguous, machine readable. Thus, we aim at creating a standard codification which could represent the basis for a new and more complete cultural appreciation of the objects studied, which will be entirely digital, as well as more accessible to a wider audience.

Introduzione

Nonostante alcune voci dissonanti sulla digitalizzazione dei testi, sembra che le *digital humanities* e, in particolare, la *filologia digitale*, stiano diventando una branca fondamentale soprattutto per lo studio dei testi letterari antichi, medievali e moderni ma non solo, perché, esse possono essere applicate anche all'iconografia in generale, alla pittura e a tutto ciò che rientra nei beni culturali².

In un'era “dell’informatizzazione sfrenata”, dove tutto deve essere conservato attraverso la “messa in rete” e dove il web, come un gigante ingordo e insaziabile, divora ogni cosa per poi restituirla alla comunità in forme e modalità differenti, anche il “sapere”, custodito nelle biblioteche o nei musei, nei manoscritti o nelle opere d’arte, ne è stato ormai risucchiato ed è, in varie configurazioni e a vari livelli, a disposizione della collettività.

Ma in che forma questo “sapere” è disponibile ed è fruibile? Esiste un ponte gnoseologico tra l’opera reale e la sua digitalizzazione che, percorrendolo, permetta al comune fruitore di capire a fondo ciò che sta guardando o sfogliando? In taluni casi sì, in altri no e spesso le informazioni lacunose e/o fuorvianti riguardano proprio le opere letterarie.

Se da un canto le nuove tecnologie hanno dato un impulso importante alla divulgazione della conoscenza rivoluzionando, in generale, i metodi di studio e di ricerca, semplificandoli, velocizzandoli e rendendoli, al contempo, ancora più attendibili perché più

completi e ricchi di informazioni, d’altro canto, purtroppo, non tutti i “prodotti scientifici” della ricerca di area umanistica sono stati registrati dalla rete. Questo mancato riversamento ha creato un divario tra l’abituale e “controllato” trasferimento delle conoscenze che avviene quasi esclusivamente attraverso la stampa, e quello incontrollato e “selvaggio” del web. Come sottolinea giustamente Francesco Stella, la facilità di pubblicare in “depositi online” o in siti non garantiti dalla presenza di un comitato scientifico, se da un lato favorisce la divulgazione della produzione culturale, dall’altro svalorizza i lavori scientificamente qualificati³. Questa situazione ha provocato un abbassamento generale del livello scientifico dei testi letterari presenti nella rete, perché non sempre adeguatamente sottoposti a un controllo qualitativo, talvolta addirittura privi di indicatori scientificamente riconosciuti, come del codice ISBN o semplicemente del DOI (*Digital object identifier*) che identifica un prodotto in ambiente digitale e che ne assicura la sua autorevolezza.

All’interno del web il grande ponte informatico che dovrebbe permettere un agevole e autorevole transito della conoscenza umanistica, e che ha il compito di rendere consapevole il fruttore di ciò che sta sfogliando o guardando, è spesso trasformato in un anonimo ponteggio traballante. E se ai fruitori, e qui ci si riferisce soprattutto ai giovani studenti, ormai disabituati nelle loro ricerche all’uso del cartaceo, non si danno gli strumenti giusti, qualitativamente elevati e scientificamente certificati, per capire a fondo un autore o un’opera letteraria, si rischia di farli cadere nella trappola dell’illusione che offre la rete che, evidentemente, non fornisce sempre notizie attendibili ed esaurienti.

In questo nuovo contesto socio-culturale la *filologia digitale* si inserisce come un’opportunità per la ricerca scientifica in campo umanistico, perché fa in modo che il trasferimento della conoscenza, attraverso il riversamento in rete di edizioni di testi letterari, possa avvenire nel modo più corretto possibile e, al contempo, che raggiunga quell’autorevolezza concessa, fino ad oggi, solo alla carta stampata.

L’obiettivo è quello di fornire alle nuove generazioni utili competenze nell’ambito dell’informatica umanistica e insegnare loro a utilizzare le nuove tecnologie per innovare e arricchire l’approccio metodologico tradizionale allo studio di un testo letterario, a condividere la propria ricerca e i risultati scientifici in modo corretto e qualificato.

La caratteristica generale di un lavoro filologico in digitale prevede la digitalizzazione di un’edizione critica di un’opera letteraria, accompagnata da un’analisi puntuale di tutte le sue componenti: delle sue caratteristiche grafiche se conservata in un manoscritto, dell’iconografia se presente e, ovviamente, della lingua, dello stile, della struttura, della versificazione se si tratta di un’opera in versi, del suo contenuto (temi affrontati, personaggi, luoghi) ecc. Uno studio filologico corredata, però, da ulteriori approfondimenti, difficilmente applicabili a un testo cartaceo, come ad esempio la creazione di un lessionario diviso per tipologia di lessico, di diverse fasce di apparato critico e di note al testo che interagiscono, attraverso un sistema di rimandi e metadati, con il testo trascritto e digitalizzato e, ovviamente, anche con le immagini del manoscritto, se presenti.

Per operare nel campo della *filologia digitale* è importante, però, acquisire determinate competenze e conoscenze senza le quali non è possibile effettuare una seria e autorevole edizione critica di un’opera letteraria, o qualunque tipo di analisi linguistica o

testuale. Malgrado il contributo tecnico che le nuove metodologie informatiche per il trattamento e lo studio dei testi conferiscono alle ricerche filologiche (digitalizzazione dei manoscritti, trascrizioni a cui applicare OCR, software per la collazione dei testi, per la ricerca di concordanze, lemmari e vocabolari online ecc), lo studio ecdotico di un componimento non può non tener conto dell'*ingenium* del filologo. Esso si sviluppa grazie al retaggio delle conoscenze acquisite durante il suo percorso formativo universitario e si affina mediante uno studio costante e continuo, e non può prescindere dalla conoscenza dei diversi metodi di critica testuale, come quello messo a punto da Karl Lachmann o da Joseph Bédier, e di tutte le successive revisioni e correzioni effettuate, da altri autorevoli filologi, nell'arco di quasi due secoli di critica letteraria. Inoltre, quando si decide di approntare un'edizione critica di un'opera in formato digitale, sarebbe auspicabile che essa avesse, comunque, anche una diffusione cartacea, malgrado i limiti che una pubblicazione a stampa impone, freni e costrizioni legate, essenzialmente, a problemi di riduzione dei costi da parte delle case editrici.

Ritornando al discorso sulla correttezza del trasferimento della conoscenza e sul profitto che la digitalizzazione di un'edizione critica di una determinata opera potrebbe apportare ai fruitori e agli studiosi, esso è legato, come sarà meglio delineato in seguito, da un lato alle metodologie utilizzate per la sua messa in rete e, dall'altro, alle infinite e svariate possibilità di studio, di analisi e di confronto offerte proprio dai nuovi software creati ad hoc per la filologia digitale.

I Un progetto di filologia digitale: edizione critica del *Romanzo di Francia* o *Libro di Fioravante*

Il progetto qui presentato è un esperimento di un'edizione critica, ancora in fieri, in formato digitale, del *Romanzo di Francia* o *Libro di Fioravante*.

Il componimento ci è stato tramandato da un *codex unicus* siglato BNF. Ital. 859, conservato alla Biblioteca Nazionale di Francia e messo in rete a gennaio di quest'anno con il titolo di *Il Fioravante*⁴. La risoluzione fotografica del codice, caricata nella sezione Gallica della BNF, è alquanto mediocre perché si tratta di immagini in bianco e nero tratte da un microfilm e non si prestano a un'edizione digitale che, invece, necessita di immagini a colori ad alta risoluzione poiché permettono una migliore lettura del codice, di analizzare la sua fattura e discernere non solo le tipologie di danneggiamento ma anche le correzioni e gli interventi nel testo effettuati nel tempo.

Il manoscritto che tramanda il *Romanzo di Francia* (RDF) è cartaceo e risale al xv secolo⁵; il testo, che occupa le cc. I-CLXXXXVII, è più antico, forse della metà del 1300. I *folii* contenenti l'opera riportano una numerazione araba più recente rispetto alla datazione del codice: il *ductus* si presenta completamente diverso rispetto a quello utilizzato dallo scriba e, inoltre, non tiene conto dei danneggiamenti subiti dal manoscritto, soprattutto della perdita di un numero impreciso di carte iniziali. Il testimone, infatti, appare abba-

stanza corrotto perché oltre ad essere acefalo è anche mutilo e, al suo interno, si ravvisano molte carte deteriorate, con strappi e macchie che, spesso, rendono illeggibile il testo.

Il *RDF* è un'opera anonima, scritta in prosa, e la lingua utilizzata è l'italiano con una forte impronta linguistica dialettale, meridionale, forse napoletana. Lo stile dell'autore è molto basso e tradisce una diffusione orale del testo che, probabilmente, solo per un concorso di circostanze favorevoli, è stato messo per iscritto ed è giunto fino a noi. Il testo è stato oggetto di una tesi di dottorato, mai pubblicata, presentata da Douglas MacArthur, dal titolo “*Il Romanzo di Francia*”. *Une version du “Libro di Fioravante” édité d’après le manuscrit unique conservé à la Bibliothèque nationale*, discussa nel 1958, e di pochi articoli nei quali, sostanzialmente, si sottolinea l'appartenenza di questo romanzo al filone letterario del *Libro di Fioravante* (xiv secolo) che, a sua volta, riprende la tradizione francese del *Roman de Floovent*⁶. Questa tradizione sarebbe stata poi contaminata da quella dei *Reali di Francia* che ingloberà nel suo contenuto anche le avventure di Fioravante.

Dall'albero genealogico ideato da Barbieri⁷, che prende in esame la tradizione manoscritta italiana del *Libro di Fioravante*, il *RDF* deriverebbe da una versione napoletana arcaica, perduta, che, a sua volta, scaturirebbe da una versione toscana contenente anche la storia di Fiovo, antenato di Fioravante, che compare anche nell'opera in oggetto. Dal punto di vista strutturale il testo si divide in due parti: la prima comincia con la narrazione delle gesta dell'imperatore Costantino ma, a causa della perdita delle carte iniziali, è possibile che iniziasse dal racconto di storie più remote; la seconda parte è interamente dedicata alle avventure di un giovane cavaliere chiamato Fioravante e del suo antenato Fiovo.

Questo lungo componimento offre ai suoi lettori, oltre a una sintesi pseudo-storica che copre un lasso di tempo che va all'incirca dal IV secolo al XIII secolo, arricchita da aneddoti e storie leggendarie, anche la possibilità di scoprire un universo linguistico molto particolare che rispecchia il parlato quotidiano del ceto sociale più basso. Al di là del contenuto, l'interesse del *RDF* risiede, quindi, nella sua veste grafico-stilistica che tradisce, come già detto, la provenienza culturale, nonché sociale, del suo anonimo autore: una persona quasi illitterata appartenente al ceto popolare. Forse un cantastorie poco acculturato, più interessato a trasmettere i fatti storici attraverso l'uso di una lingua più vicina al suo potenziale ascoltatore illitterato, che intenzionato ad acculturare il suo pubblico mediante il ricorso a una forma espressiva corretta e raffinata che, evidentemente, non sarebbe stata compresa, né apprezzata dal suo uditorio.

Espressioni e frasi gergali come: *ca mo mo; ca ponno; ca ve avite fatto; foro issiute, dare la ciccia, gitato a rame, uochie de la menti* ecc.; lemmi tipicamente dialettali come: *battiare* per “battezzare”; *igranare* per “ingannare”; *isbrandente* per “splendente”; *groria frama* per “orifiamma” e ancora *visigna, rebruso, lamita, ciciate, golio, giosse, citella, aratrare, ciorfetara, vanda, sunescarco, namorata, denochia*, nomi di luoghi e personaggi storici graficamente travisati: *Iellusalem, Sudano di Babellia, Igreterra, Berlino* per Merlino, *Utter Pandraone* per Uter Pendragon, evidenziano la peculiarità linguistica dell'opera. Per lo studio linguistico dell'opera sarà necessario analizzare tutta la produzione letteraria dialettale dell'epoca e in modo particolare i *Ricordi* di Loise de Rosa, un importante testo

napoletano datato da Formentin tra la fine del XIV e inizio XV secolo⁸ che, da una prima ma ancora superficiale indagine, sembra registrare dei tratti grafico-fonetici molto simili a quelli registrati nel *RDF*. Ovviamente si tratta solo di un'ipotesi che solo una rigorosa e accurata analisi linguistica del testo potrà confermare o confutare⁹.

Senza dubbio gli aspetti linguistici del *RDF*, uniti a una narrazione stilisticamente altalenante nella quale si passa improvvisamente dal racconto impersonale, al dialogo, all'uso della prima persona e poi della terza, come se il testo fosse il risultato di una serie di appunti presi durante una recitazione o, comunque, di una trasmissione orale dei vari racconti e storie che compongono il testo, fanno di quest'opera un *unicum* nella storia letteraria dialettale italiana del XIV secolo.

Già da questa sintetica presentazione si evince che il *RDF* offre tanti spunti di riflessione e di ricerca (la struttura linguistica, il lessico particolareggiato, lo stile, gli eventi storici narrati e quelli leggendari, i personaggi talvolta reali e altre volte inventati e altro ancora), e quindi si tratta di un testo che si presta bene a un'edizione in formato digitale. Data la lunghezza dell'opera, l'edizione digitale verrà effettuata in diverse fasi: si procederà prima con una trascrizione diplomatico-interpretativa del componimento e poi, man mano, si aggiungeranno gli apparati e le diverse tipologie di analisi testuale. Tramite l'ausilio degli strumenti informatici, oltre alla realizzazione di un sistema di rimandi tra la trascrizione, la traduzione dell'opera e le immagini del codice, sarebbe, difatti, anche possibile creare un Thesaurus articolato che contempli varie tipologie lessicali dialettali: locuzioni, proverbi, lemmi, verbi, nomi di luoghi e dei personaggi. Inoltre, per ogni avvenimento narrato, potrebbero essere creati degli approfondimenti storici e letterari, arricchiti da rinvii alle altre opere che presentano le stesse tematiche e, ovviamente, corredare il testo con un'introduzione all'edizione, uno studio linguistico-dialettale, un apparato critico e delle note al testo. Si tratterebbe quindi di creare una classica edizione critica che, grazie all'interazione con gli strumenti informatici, permetterebbe di effettuare tutta una serie di analisi e indagini scientifiche difficilmente contemplabili da un testo a stampa.

2

Filologia e digitale: quali ricadute

Lo studio di materiali testuali con l'ausilio del computer ha dato vita alla definizione di standard e linguaggi per la costituzione di corpora. Lo standard TEI di descrizione dei documenti e i linguaggi SGML e XML sono un concreto esempio dello sviluppo trasversale di applicazioni software per la decodifica dei testi, ad uso degli umanisti¹⁰.

Lo studioso che si appresta a descrivere/codificare un testo antico¹¹, oggi come ieri, si trova sostanzialmente a dover compiere delle scelte riguardanti le informazioni da conservare. È dunque utile partire dalla concreta conoscenza della storia della filologia, con particolare riguardo per il periodo medievale e umanistico affinché sia possibile entrare nel merito dei metodi e delle scelte dei filologi. Tutto ciò deve poi trovare una

sua concreta applicazione nello studio degli standard di descrizione bibliografica¹² che, negli specifici casi di creazione di data base relazionali¹³, rappresentano un sicuro e gestibile punto di incontro, ormai solido, tra le buone pratiche della biblioteconomia e le esigenze dell'informatica. Questo per garantire un prodotto versatile a livello software pur mantenendo inalterate le consolidate pratiche della filologia, della bibliologia e della descrizione bibliografica.

La codifica diventa, quindi, un atto interpretativo attraverso cui evidenziare, da una parte, gli aspetti strutturali quali ad esempio segnatura, impronta, spazi ecc., dall'altra il significato intrinseco del segno, manoscritto o a stampa.

L'oggetto libro è il frutto dell'ingegno umano nel suo contenuto, il testo trasmesso/pubblicato, e nel suo contenitore (supporto, decorazione, legatura) e va dunque trattato con il medesimo riguardo delle altre opere d'arte¹⁴.

Il filologo e il bibliotecario conservatore sono i primi attori nel campo dello studio, valorizzazione e capacità di messa in fruizione del testo e del contesto legati all'oggetto libro; vanno per questo curati alcuni aspetti didattici, a volte poco considerati che, nell'epoca di Internet, legano indissolubilmente la capacità di descrivere in maniera esaustiva un volume e l'attività di decodificazione del contenuto da inserire in un *data base*.

È dunque importante, a mio avviso, per entrambe le figure professionali, partire dalla conoscenza degli Standard per la catalogazione dei manoscritti delle biblioteche italiane¹⁵ e degli ISBD(A)¹⁶ per il libro antico attraverso i quali vengono compilate le descrizioni tipiche degli OPAC (catalogazione partecipata) e vengono anche gestite le informazioni a livello esemplare (postille manoscritte, interpolazioni del testo, *ex libris*, provenienze ecc.) oltre all'allestimento dei campi del Dublin Core del MAG.

Da questo discorso non possono essere omessi i formati SGML (*Standard Generalized Markup Language*) e l'XML (*Extensible Markup Language*) attraverso cui è possibile dichiarare la funzione di ogni specifico elemento testuale¹⁷ preparando adeguati metadati (secondo gli approfondimenti richiesti).

Nel caso delle edizioni critiche da gestire in web, vanno inoltre pensati testi e tesauri multilingua all'interno dei quali le liste di autorità possano essere gestite in almeno due lingue, per favorire le ricerche e diffondere al meglio il lavoro svolto; si rendono anche indispensabili liste codificate di argomenti/soggetti a corollario dell'edizione fruenda in rete che possano richiamare attraverso i grafi elementi simili disseminati nel web (collegamenti ad altri oggetti digitali). Tutti strumenti utili a supporto dell'edizione critica in formato digitale.

Naturalmente, prima di arrivare all'edizione digitale, la fruizione di codici manoscritti, volumi antichi, documenti, passa attraverso la ricerca in biblioteca o in archivio (mi reco nell'istituto culturale e studio in loco ciò che mi interessa) oppure, in maniera più veloce e diretta, ove fossero già presenti *data base* o teche digitali, attraverso la consultazione dei preziosi manufatti messi in rete e dunque disponibili all'utenza interessata, attraverso Internet. A seguito dell'individuazione dell'opera da studiare, analizzare, collazionare ed editare, si intraprende un percorso funzionale alla valorizzazione e fruizione in edizione critica di un testo; è dunque opportuno ricordare in

maniera obiettiva, i vantaggi e i limiti dell’acquisizione digitale praticata ormai da oltre un decennio dagli istituti culturali.

In primis è necessario porsi delle domande: dal punto di vista prettamente filologico, quanto l’acquisizione ottica può influire in termini di semplificazione delle procedure, diminuzione del tempo di confronto tra esemplari o testimoni (in filologia)? Qual è il giusto rapporto tra costi e benefici, quale l’impatto sulle attività di ricerca scientifica e fruibilità dei dati? Esiste uno schema di valutazione d’impatto di simili progetti/attività?

Indubbiamente la digitalizzazione offre la possibilità di gestire più livelli di lavoro. Tra i vantaggi materiali relativi alle comodità/funzioni offerte dal digitale vanno sottolineati i seguenti aspetti:

- riproduzione a costo minore¹⁸;
- unione testo/immagine + altri oggetti digitali/digitalizzati;
- massima diffusione sul Web¹⁹;
- inclusione di materiale linkabile da altri siti o risorse in rete;
- navigazione ipertestuale (se prevista, ove possibile);
- strumenti di ricerca (produzione liste e concordanze, software per il restauro digitale, liste di soggetto funzionali al reperimento dei dati o alla sintetizzazione di argomenti ecc.).

Tra i limiti che l’uso del digitale comporta, vanno per chiarezza enunciati: primo fra tutti la difficoltà di produzione/riproduzione dei supporti che, quando non sono scaricabili online, vanno ugualmente richiesti alle Biblioteche di riferimento le quali, per una riproduzione digitale, possono chiedere cifre molto alte. Non va inoltre trascurato l’aspetto più tecnico ma ugualmente importante della frammentazione a livello software (configurazione di piattaforme digitali ad hoc, requisiti di sistema, longevità e conservazione a lungo termine variano enormemente a seconda delle scelte operate dal team di progetto).

Ultimo punto, ma non meno importante, è tener conto del rapporto con l’utente finale che comunque usufruirà di un prodotto digitale visionabile comodamente dal proprio personal computer. È dunque utile calcolare al meglio il livello dell’interfaccia utente, affinché l’utilizzatore del sistema possa avere ben chiari i parametri di interazione con esso: posso intervenire sul testo, annotarlo? Esiste una *chat* o una *community* con la quale discutere? Posso consultare i testimoni utilizzati? Posso vedere/raggiungere altri oggetti simili o in relazione con quanto ho trovato nel contenitore digitale? Ho una visione diretta di eventuali bibliografie correlate o collegamenti ad altri siti/DB *open source*? Domande che meritano un approfondimento dal punto di vista della valutazione del progetto editoriale a breve, medio e lungo termine.

Si dimostra utile quindi un approccio didattico non limitativo: studenti di corsi di laurea magistrale, laureandi, dottorandi, ricercatori, interessati a un’introduzione ai temi delle *digital humanities*, devono poter includere nella loro esperienza di tipo tradizionale, anche un approccio “pratico”, con sessioni di lavoro, lezioni e *workshop* dedicati alla preparazione e visualizzazione di materiali digitali. Quello che nell’ambito delle

biblioteche viene concepito come il *semantic-web*²⁰ può essere senz’altro un utile punto di partenza e acquisizione di conoscenze utili allo scopo!

E ancora, testi medievali manoscritti e testi di iscrizioni antiche sono strettamente legati a nuove ricerche sui sistemi di scrittura e dunque allo sviluppo di programmi a supporto delle investigazioni epigrafiche o finalizzati alle indagini sulla distribuzione dei segni, attraverso i secoli e le diverse culture. Progetti di OCR per lingue antiche o tesauri sono stati sviluppati dal CNR²¹ e anche presso l’Università degli Studi di Salerno in seno al Laboratorio DOC [Dizionario online di OCCitano] di Filologia romanza. Il quadro generale dimostra quindi come alcune scienze, la filologia digitale è fra queste, stiano aprendo a nuove generazioni di studiosi, strumenti, applicazioni e tecnologie che favoriscono un’interazione dinamica tra discipline e saperi. La prima fucina è rappresentata dalla didattica curriculare e dalle successive esperienze laboratoriali: studio della produzione testuale e analisi semantica e sintattica, costruzione di archivi digitali, riflessioni teoriche e metodologiche sulle nuove possibilità offerte dai software per le edizioni critiche digitali²², utili ad esempio, per segnalare più passi del testo contemporaneamente (con l’ausilio di evidenziatori elettronici), così da individuare in modo preciso il “luogo” in cui si trova una parola, una occorrenza, un passo. Da qui la possibilità di creare indici normalizzati dai quali attingere per lo studio delle forme verbali, delle assonanze del lessico in generale, fino alla scelta delle varianti nei casi di tradizioni con più testimoni con diverse corrispondenze. Dunque il significato della parola può essere approfondito studiando i passi in cui essa ricorre; e i passi a loro volta, pur nella loro diversità, risultano in relazione tra loro dall’uso di quella medesima parola. Un tale approccio al contenuto dell’opera traddita, informatizzato, potrebbe aumentare esponenzialmente le possibilità di navigazione infratestuale, permettendo di collocare i termini in una determinata tipologia testuale²³.

Tutto parte dalla concreta applicazione degli standard citati, dall’avanzare delle tecnologie dei *linked open data*²⁴ e dei *big data*²⁵ come strumenti di raccordo e condivisione delle informazioni.

Ecco perché vecchie e nuove figure professionali si intrecciano:

- bibliotecario e archivista “digitale”;
- filologo digitale;
- *digital curator*;
- consulente digitale;
- *big data specialist*²⁶.

Dal punto di vista della pubblicazione del testo digitale, prendono vita forme che affiancano e determinano il nuovo processo editoriale; si tratta di figure chiave che interagiscono fra loro (e spesso si sovrappongono), dalle quali, come si può notare, dobbiamo assolutamente pretendere competenze trasversali, come ad esempio le capacità applicative relative a filologia, lingue, bibliologia e bibliografia che devono essere ben affiancate da abilità tecniche, uso di specifici software e una base di informatica. Un panorama che vuole definirsi professionalizzante a più livelli, induce dunque allo studio e alla comprensione di più chiavi d’accesso al sapere; il classico “chi fa cosa?” si traduce in un intreccio di competenze necessarie alla buona riuscita di qualsiasi ricerca. Per fare

un esempio, non certo esaustivo, di reciprocità tra attività lavorative, tipico dei nuovi panorami multidisciplinari in cui si muovono i beni culturali²⁷, possiamo ragionare su alcuni elementi e figure professionali:

- Autore del progetto (bibliotecario – archivista-*digital curator-big data specialist* – informatico): definisce il progetto dal punto di vista scientifico e ne pianifica le metodologie di applicazione informativa e informatica ed è responsabile della sua realizzazione.
- Curatore-filologo/redattore (filologo digitale – consulente digitale – *digital curator*): scrive e/o edita i testi, collaziona i testimoni, predisponde la genealogia del testo, ne studia la lingua e ne riconosce la derivazione geografica; imposta e integra l’edizione digitale.
- Codificatore/decodificatore (bibliotecario – filologo – archivista-*digital curator*): descrive i testi attraverso i protocolli internazionali, aggiunge ai testi il *markup*, li codifica attraverso i metadati, tutto secondo gli standard di riferimento.
- Grafico (informatico–consulente digitale – *digital curator*): elabora il progetto grafico seguendo i protocolli e gli standard di settore.
- Programmatore (informatico – consulente digitale – *digital curator*): progetta, realizza e cura il software, la piattaforma, carica immagini e metadati.

L’esempio dimostra come gli elementi che contraddistinguono la nascita e pubblicazione di una edizione critica digitale, non possono essere seguiti solo da informatici (anche se specializzati), parliamo infatti di processi di edizione dei testi in cui la “correttezza” è assicurata dallo specialista, in questo caso il bibliotecario, il filologo ecc. con il suo *knowhow* umanistico²⁸.

3

Il progetto di edizione critica digitale: il cantiere

La storia della navigazione e fruizione in Internet di oggetti digitali è piuttosto ricca. Negli anni Ottanta dello scorso secolo, le università americane hanno iniziato ad interconnettersi attraverso la rete: il primo gruppo di discussione di ambito umanistico fondato da Patrick Conner nasce nel 1986, seguito da Humanist, il *newsgroup* dedicato ai temi delle *digital humanities and humanities computing*.

Con l’apparire dello *Standard Generalized Markup Language* (SGML)²⁹ come linguaggio ISO per gestire vari tipi di testo, si comincia ad avvertire il bisogno di uno standard di marcatura dedicato all’ambito umanistico. Successivamente con la fondazione del *World Wide Web Consortium* (1994), vengono pubblicate le prime TEI *guidelines* e nel 1999 viene fondato il TEI *Consortium*³⁰.

Il quadro generale si declina in un numero elevato di tecnologie, linguaggi di programmazione e marcatura, strutture di dati, protocolli di rete con indiscutibili ricadute interdisciplinari. Queste componenti concorrono a realizzare quello che è il prodotto finale, l’oggetto informativo digitale, sia esso disponibile attraverso la rete, un CD o un software.

La scelta di quali componenti utilizzare in funzione del risultato che si vuole ottenere non è né semplice né automatica, verrebbe da dire “meccanica”, e presuppone tutta una serie di considerazioni, sia qualitative sia quantitative, in cui un corretto giudizio, verrebbe da dire “critico”, è fondamentale³¹.

L’edizione critica in formato digitale ha a sua volta caratteristiche peculiari: il filologo può scegliere in base all’oggetto della ricerca o alla quantità dei testimoni; può sviluppare un’edizione diplomatica, o interpretativa, o mista, in formato elettronico, da distribuire su supporto digitale o via web. L’edizione potrà essere *web-based* solo testo, *stand-alone* testo e immagini (uso di un navigatore web software specifico) oppure si può optare per una *full digital edition* che contempli il testo dell’edizione, tutte o buona parte delle varianti e le immagini del manoscritto.

Va ricordato che al momento sono stati creati soprattutto strumenti per effettuare la collazione in modo semi-automatico (è impossibile, per ora, una collazione del tutto automatica).

L’importante è imparare a conoscere e usare lo strumento digitale senza tralasciare di partire sempre dalle competenze di tipo tradizionale che, non dimentichiamo, sono parte integrante delle attività didattiche e formative caratterizzanti i percorsi umanistici. Non bisogna dimenticare che il metodo che introduce alla filologia digitale è migliorativo e non sostitutivo dell’approccio teorico-pratico classico che resta solidamente alla base di qualsiasi intervento in questo senso.

Il digitale offre quindi una serie di possibilità che passano attraverso una buona struttura d’insieme; esso deve sintetizzare almeno tre buone pratiche:

- coerenza del metodo filologico utilizzato;
- sicurezza e riproducibilità dei dati immessi;
- garantire che il lavoro sia “usabile” e “conservabile” a lungo termine.

All’apertura del “cantiere” di lavoro abbiamo dunque dovuto ragionare sugli aspetti appena sintetizzati optando per la semplicità. Inizialmente abbiamo lavorato sugli aspetti relativi alla lingua, allo stile, agli eventi narrati, senza però trascurare il supporto.

Un primo passo decisivo a livello tecnico è stato l’individuazione di un software adatto allo scopo e sufficientemente snello per l’interrogazione del corpus e delle eventuali varianti significative. Nello spirito dell’utilizzo di risorse *open source*, abbiamo quindi pensato di scaricare qualcosa dalla rete, partendo dall’assunto che, un buon software, adatto alle esigenze di un filologo con competenze medie di informatica, deve avere una interfaccia piacevole con icone parlanti e possibilmente moduli attraverso i quali sia possibile una navigazione amichevole. Abbiamo quindi scelto Juxta con il quale è possibile effettuare una collazione semi automatica³².

Lo stato dell’arte: il futuro della filologia digitale si gioca sulle strategie che consentono di scambiare, riutilizzare e reinventare materiali, così tra le priorità del progetto è annoverata senz’altro la capacità di trasformazione della base dati attualmente progettata e codificata in XML³³ traducibile in caso di necessità in altri formati. Questo perché la parte relativa alla gestione del testo con Juxta³⁴ è solo un ambito

di un progetto più ampio. Il sistema di indicizzazione e algoritmi per il recupero dei dati derivano dall'ambito biblioteconomico, molto avanti nella strutturazione di *data base* che gestiscono informazioni bibliografiche con immagini e metadati. Si tratta di un sistema completamente nuovo all'interno del quale, la trascrizione e gestione del testo critico è parte di un complesso organico di contenitori digitali, che ha una sua struttura formale dedicata.

Per queste ragioni un cambiamento del sistema di marcatura del testo base deve essere sempre possibile, per restare al passo con i mutamenti del web, e rendere concretamente flessibile e sempre aggiornabile il sistema anche se una operazione del genere richiede la riscrittura del *parser* automatico (input, dei dati/elementi letti per esempio da un file o da una tastiera) e di conseguenza la riconfigurazione del motore di ricerca, naturalmente da effettuarsi anche in questo caso, secondo criteri ascrivibili a standard ben strutturati, capaci di assicurare stabilità, efficienza, efficacia e velocità³⁵.

Nel contenitore denominato Thesaurus/Lemmario posto al centro dell'immagine (FIG. 1), si prevedono una serie di *box* virtuali per accogliere le particolarità del testo: forme verbali, luoghi geografici, forme linguistiche, nomi di persona normalizzati con le varianti a costituire il blocco semantico relativo all'*Authority file*.

Di seguito alcune esemplificazioni tratte dagli elenchi che stiamo compilando estraendo i lemmi dal testo. Tra le forme verbali: *facenno, foy, partìò, vinesene*; dalla lista riferita ai luoghi: *Avingnone, Millana, Parisse, Provencia*. Interessante l'onomastica, con personaggi come l'imperatore Costantino, papa Sabiniano, san Dionigi ecc.

Di grande importanza per uno studio complessivo della lingua del manoscritto, sono anche alcune particolarità linguistiche e lessicali dalle quali si ricavano considerazioni sull'origine del testo: *battallia, battiare, mo mo, trase... ecc.*

Su un paio di queste liste, nomi di luoghi e di persona, bisognerà tornare per effettuare una normalizzazione delle forme in modo da creare un thesaurus coerente, esportabile e funzionale alla configurazione di eventuali ricerche ipertestuali e per i metadati.

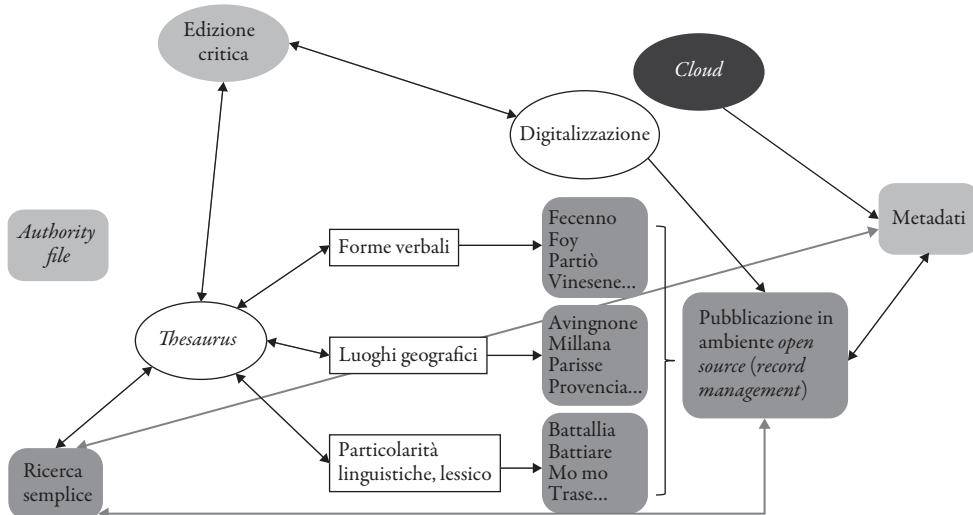
Il progetto prevede, come si è già accennato, l'utilizzo degli strumenti della biblioteconomia, e, nello specifico, della descrizione catalografica dei testi. La normalizzazione si basa sul soggettario della BNCF³⁶ e sulle REICAT³⁷ così che i lemmi siamo ben riconoscibili a livello HW e SW nei sistemi informativi.

Si intende gestire con la medesima attenzione anche le varianti dei nomi, luoghi, personaggi, per popolare il DB con il maggior numero di "soggetti" possibile. Le entità, integrate nello spazio virtuale, conserveranno un ordine gerarchico che mette al primo posto, ad esempio per i nomi, l'intestazione riconosciuta da REICAT per poi comprendere tutte le altre forme. La pianificazione di un *Authority file* multilingue ci sembra poi la migliore soluzione per venire incontro a una platea di studiosi collegati alla rete!

Di seguito alcuni esempi di intestazione normalizzata (TABB. 1 e 2: luoghi e nomi).

Il caso di studio presentato dimostra come la pratica dell'edizione digitale spinga a riconfigurare gli obiettivi del lavoro filologico, in chiave maggiormente indagatrice an-

FIGURA I
Esempio di thesaurus in DB



che rispetto al lavoro editoriale, alla capacità/possibilità di interazione con gli elementi informatici e con le pratiche e gli assunti teorici e metodologici più innovativi.

Non si può sottovalutare l’impatto che il digitale sta avendo sulle pratiche editoriali; i cambiamenti che si stanno producendo sono infatti profondi e riguardano non solo il metodo di lavoro dell’editore (cioè l’euristica editoriale), ma investono anche gli obiettivi scientifici dell’edizione³⁸.

L’auspicio, dal punto di vista strettamente filologico e, nel contempo, punto di incontro tra le scienze del libro e quelle del testo, è di giungere alla realizzazione di uno “stemma codicum” evoluto, che mi piace definire “parlante”, attraverso il quale si possa facilmente navigare, oltre che attraverso le segnature e le localizzazioni dei codici oggetto di studio, anche trasversalmente (descrizione bibliologico-bibliografica, digitalizzazione e trascrizione dell’opera); con un simile risultato, il lavoro sui singoli testimoni, con le varianti e l’apparato, alla base delle procedure di collazione finalizzate a provare le scelte del filologo per l’edizione critica, potranno essere sempre consultabili.

Conformare una collezione digitale al linguaggio di *markup TEI* (di edizioni scientifiche) con l’ausilio delle “scienze sorelle”, diventa un imperativo necessario per riuscire a far parte della struttura internazionale legata alle iniziative in *digital philology*.

TABELLA 1
Esempi di intestazione normalizzata – Luoghi

Luogo, termine a testo	Luogo normalizzato (fonte SBN)	Varianti (fonte SBN)	In altre lingue francese(fr); inglese(en); tedesco(de)
Avingnone	FR Avignone	Auenione; Avignon; Auenioni...	Avignon(de)(fr); Avignone(en)
Millana	IT Milano	Melano; Mediolani; Milan ...	Milano(de)(fr) (en)
Parisse	FR Parigi	Parisjis; Lutetiae Pariso- rum; Parisius; Paris...	Paris(de)(en) (fr)

TABELLA 2
Esempi di intestazione normalizzata – Nomi

Nomi di persona: termine a testo	Nome normalizzato (fonte SBN)	Varianti (fonte SBN) – CERL (Thesaurus)	In altre lingue francese(fr); inglese(en); tedesco(de) ¹⁸
Costantin	Constantinus <im- peratore; 2.>	Flavius Claudius Con- stantinus <imperatore>; Costantino <imperatore; 2.>	Konstantin <I., Rö- misches Reich, Kai- ser>(de); Constantine I, Emperor of Rome, – 337(en); Constantin I (empereur romain; o272-o337)
Savignano	Sabinianus <papa; sec. 7.>	Sabinian <Papst> Sabinian <von Volterra> Sabiniano <Papa> Sabiniano <di Blera> Sabinianus <de Volterra> Sabinien <Pape> Sabinien <de Volterra>	Sabinianus, Papa(de); Sabinianus (pape; – 606) (fr); Sabinian, Pope (en)
San Donisse	Dionigi <santo>	Dionigi di Parigi	Dionysius von Paris (de); Dionysius, von Paris(fr); Dionysius, Saint approxi- mately dead 250-285(en)

Conclusioni

Se da un lato un’edizione digitale di un testo prevede, come già più volte sottolineato, una buona competenza sia filologica che informatica, le ricadute che essa assicura in ambito scientifico, didattico e professionale, ripagano del complesso lavoro effettuato.

Dal punto di vista scientifico un'edizione in formato digitale ha una circolazione più vasta rispetto a un'edizione a stampa e, sicuramente, più rapida. Inoltre essa permette un'apertura e un'interazione tra studiosi appartenenti a vari settori scientifici di ambito umanistico, anche se molto distanti tra loro geograficamente, che, semmai, stanno effettuando studi analoghi: come ad esempio bibliofili e paleografi (con la presenza della digitalizzazione dell'intero codice), linguisti (con la creazione di un Thesaurus commentato e di un sistema di concordanze linguistiche), letterati (con la presenza di un apparato critico e di note al testo), storici (con la presenza di un'analisi e di un commento agli eventi o ai personaggi storici presenti nel testo), studiosi di discipline artistiche (con la presenza e descrizione di eventuali miniature presenti nel codice e corredi artistici). Infine, ed è l'obiettivo più importante, si porterebbe in rete un lavoro "qualificato", di alto valore scientifico, che andrebbe a offuscare tutto ciò che, nel *web*, non è contraddistinto come tale.

Dal punto di vista didattico le ricadute di un lavoro del genere sono infinite e, ovviamente, diversificate per gradi di complessità, di tipologia di studenti e di utilità nell'ambito del corso o del percorso formativo che un docente vuole affrontare. Al di là di tutto, una buona edizione critica in formato digitale di un qualsiasi testo è di grande ausilio per presentare e analizzare un'opera letteraria in tutti i suoi aspetti e peculiarità e, senza dubbio, rende le lezioni e/o lo studio ancor più coinvolgenti e partecipativi.

Dal punto di vista professionale si ribadisce la necessità, che sta emergendo soprattutto in Italia perché ancora poco attiva nel campo della filologia digitale rispetto ad altri paesi europei, di trasmettere ai più giovani queste nuove metodologie e, quindi, di insegnare agli studenti, ai laureati, ai dottorandi, l'uso di software per l'informatica umanistica e ad iniziargli all'apprendimento delle norme, come ad esempio TEI, per strutturare un progetto di edizione di digitale. L'utilizzo degli standard e la conoscenza trasversale di materie affini, funzionali alla formazione di nuove figure professionali sempre più versatili, è una missione che i centri di formazione come le università non possono ignorare. L'impiego dei saperi della biblioteconomia, i descrittori internazionali e i formati informatici per l'allestimento dei *data base* relazionali con l'ausilio dei *linked open data* e le funzionalità offerte dai grafi, possono plasmare in positivo il futuro delle nuove generazioni di *digital curator*, consulenti digitali, filologi e bibliotecari digitali. Questo apprendimento permetterà loro, da un lato, di continuare ad approfondire le conoscenze in ambito letterario, linguistico e filologico e, dall'altro, di acquisire specifiche competenze nelle *digital humanities*, da utilizzare in vari contesti lavorativi, dove sono e saranno sempre più richieste figure specializzate in questo campo. In maniera sempre più incisiva, in molti settori lavorativi, il saper applicare l'informatica alle scienze umane sta diventando un requisito professionale indispensabile e questa è una realtà che l'area umanistica del mondo accademico non può più ignorare.

Oggi come ieri, l'obiettivo principale di chi fa ricerca è trasmettere e tramandare non solo i risultati dei propri studi ma il "Sapere" in generale e, per far sì che esso diventi largamente condivisibile, è necessario seguire l'evoluzione dei tempi e dei mezzi di trasmissione della conoscenza. Questo non vuol dire cambiare le metodologie tra-

dizionali, non vuol dire, riprendendo le parole di Tomasin, sostituire con un approccio riduttivamente tecnico «l’impiego del miglior hardware di cui disponiamo, cioè il nostro cervello»³⁹, significa semplicemente “aprirsi” ad altre e nuove possibilità affinché le edizioni e gli studi scientifici di qualità, possano essere raggiungibili da una sempre più ampia comunità di lettori.

Note

1. Sabrina Galano è autrice dell’Introduzione, del paragrafo 1 e delle Conclusioni; Maria Senatore Polisetti è autrice dei paragrafi 2 e 3.

2. Gli studi e i dibattiti sull’informatica umanistica sono in continua evoluzione e la bibliografia è molto vasta (si vedano in particolare i contributi di L. Spinazzé, *Filologia digitale. Dalla ricerca alla didattica*, Tangram Edizioni Scientifiche, Trento 2015 e L. Tomasin, *L’impronta digitale: cultura umanistica e tecnologia*, Carocci, Roma 2017 che, pur apprezzando l’utilità delle nuove tecnologie, critica le metodologie di edizione digitale che non rispettano e, spesso, stravolgono il tradizionale approccio filologico all’analisi testuale). Francesco Stella ha pubblicato di recente una breve “guida pratica” alla scelta e all’uso di strumenti digitali (software e programmi) realizzati negli ultimi anni, per l’analisi e l’edizione dei testi, corredata da una corposa bibliografia e da un elenco di siti web (piattaforme, biblioteche, edizioni digitali) molto interessanti e utili per chi volesse intraprendere un progetto di edizione digitale (F. Stella, *Testi letterari e analisi digitale*, Carocci, Roma 2018).

3. *Ivi*, p. 17.

4. Si veda il sito ufficiale <https://gallica.bnf.fr/ark:/12148/btv1b10033064w/f5.image> (ultimo accesso 12 febbraio 2019).

5. Cfr. G. Mazzatinti, *Inventario dei manoscritti italiani delle biblioteche di Francia*, Presso i principali librai, Roma, 1887, pp. 217-26.

6. Cfr. D. MacArthur, *Les versions du Libro de Fioravante*, in “Filologia Romanza”, VII, 1-2, 1960, pp. 121-8 e N. I. Barbieri, *Le versioni italiane della storia di Fioravante*, in “Rendiconti di Lettere”, 145, 2011, pp. 107-26.

7. *Ivi*, p. 126.

8. V. Formentin, *Loise de Rosa “Ricordi”. Edizione critica del ms. ital. 913 della Bibliothèque nationale de France*, tomo 1, Salerno Editrice, Roma 1998.

9. Uno studio più approfondito sugli aspetti stilistici e linguistici del testo è in corso e sarà oggetto di una successiva pubblicazione.

10. Medesime tematiche sono state affrontate già da Gigliozzi, Fiormonte, Numerico e Vespiagnani nel 2003; G. Gigliozzi, *Introduzione all’uso del computer negli studi letterari*, a cura di F. Ciotti, Bruno Mondadori, Milano 2003; D. Fiormonte, *Scrittura e filologia nell’era digitale*, Bollati Boringhieri, Torino 2003; Teresa Numerico, Arturo Vespiagnani (a cura di), *Informatica per le scienze umanistiche*, il Mulino, Bologna 2003.

11. Numerose sono le pubblicazioni sulla codifica e decodifica dei testi e sulla descrizione bibliografica di materiali antichi e moderni. Per citare solo alcuni contributi relativi agli aspetti di gestione delle informazioni a livello informatico: E. Pierazzo, *La codifica dei testi: un’introduzione*, Carocci, Roma 2005; L. Burnard, C. M. Sperberg-McQueen, *Il manuale TEI Lite: introduzione alla codifica elettronica dei testi letterari*, a cura di F. Ciotti, Sylvestre Bonnard, Milano 2005; S. Maffei (a cura di), *XML per i beni culturali: esperienze e prospettive per il trattamento di dati strutturati e semistrutturati*, Edizioni della Normale, Pisa 2007; M. Tettamanzi, *Html5: guida tascabile al linguaggio e agli elementi di una pagina web*, Apogeo, Milano 2017; F. Colace et al. (eds.), *Data Management in Pervasive Systems*, Springer, New York 2015.

12. Qui si intendono gli ISBD, le Reicat e i sistemi di marcatura dei software di catalogazione: Unimarc e Marc21 in primis.

13. A partire dai principi evidenziati dalle regole di Codd. Si veda: E. F. Codd, *The Relational Model for Database Management: Version 2.*, Addison-Wesley, Reading (MA) 1991. Si veda anche, per un approccio semplificato al tema: https://it.wikipedia.org/wiki/12_regole_di_Codd (ultimo accesso 12 febbraio 2019).

14. Si veda, tra gli altri: G. Di Domenico, G. Paoloni, A. Petrucciani (a cura di), *Percorsi e luoghi della conoscenza: dialogando con Giovanni Solimine su biblioteche, lettura e società*, Bibliografica, Milano 2016; E. Gennaro (a cura di), *Biblioteche musei archivi: quali sinergie?*, Provincia, Ravenna 2012.

15. Si veda l'ultima edizione disponibile in rete: L. Merolla, L. Negrini (a cura di), *Guida a ManusOnLine (MOL) Standard per la catalogazione dei manoscritti delle biblioteche italiane*, ICCU, Roma 2014, https://manus.iccu.sbn.it/GUIDA_settembre_2014.pdf. (ultimo accesso 12 febbraio 2019).

16. ISBD (A): *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)*, International Federation of Library Associations and Institutions; recommended by the ISBD Review Group of the IFLA Cataloguing Section. [S. l.: s. n.], 2006.

17. SGML è un meta-linguaggio, un insieme di regole usate per creare linguaggi speciali che prendono il nome di *markup language*. Le applicazioni più note di SGML sono HTML, TIM (Telecommunication Interchange Markup). È largamente usato come strumento di immagazzinamento e scambio di informazioni (dati). Si usa anche in tipografia. XML eredita da SGML la capacità di definire in maniera semplice nuovi marcatori, creando di fatto dei linguaggi di *markup* personalizzati, mentre la complessità e le caratteristiche opzionali che appesantivano l'SGML sono state eliminate dall'XML. Si veda a titolo esemplificativo: C. F. Goldfarb, *The SGML Handbook*, edited and with a foreword by Yuri Rubinsky, Clarendon Press, Oxford 1990; F. Cusimano, *Due esempi di buone pratiche nell'uso dei metadati XML: un'efficace disseminazione dei contenuti digitalizzati*, CRELEB – Università cattolica, CUSI, Milano 2014.

18. Di norma maggiore è la quantità delle immagini da digitalizzare, minore è il costo per singolo file. Un costo che tutela committente e fornitore si dovrebbe aggirare tra i 0,20 e i 0,30 centesimi a immagine (su medie grandi quantità). Esistono sul mercato prezzi anche più bassi per lavori massivi; il consiglio è sempre quello di badare alla qualità e alla metodologia di rilascio dei file conservativi (l'hard disk esterno consegnato alla committenza con i metadati resta ancora il miglior sistema di consegna. Le immagini senza i metadati non sono utilizzabili da sole! Tutto in *cloud* è comodo ma potrebbe essere poco funzionale per chi verrà dopo di noi!)

19. È utile pensare di valutare l'impatto e la qualità del progetto a medio e lungo termine, attraverso i dati che è possibile raccogliere dalla rete (visibilità, fruizione, possibilità di interazione, velocità di scarico dei file ecc.). La valutazione d'impatto in questi casi va pensata in prospettiva triennale.

20. Il web semanticò è stato ideato da Tim Berners-Lee. Si intende la trasformazione del World Wide Web in un ambiente dove i documenti pubblicati (pagine HTML, file, immagini, e così via) sono associati ad informazioni e dati (metadati) che ne specificano il contesto semantico in un formato adatto all'interrogazione e l'interpretazione (es. tramite motori di ricerca) e, più in generale, all'elaborazione automatica. Per un approccio semplice al tema: https://it.wikipedia.org/wiki/Web_semanticò (ultimo accesso 12 febbraio 2019). Per approfondire il tema si vedano, tra gli altri: J. Powell, M. Hopkins, *A Librarian's Guide to Graphs, Data and the Semantic Web*, Chandos, Amsterdam 2015; G. Dunsire, *RDA and the Semantic Web: Lectio magistralis in Biblioteconomia: Firenze*, Università degli studi di Firenze, 4 marzo 2014, Casalini Libri, Fiesole 2014; T. Di Nola et al. (a cura di), *Semantic Web: tra ontologie e Open Data*, Apogeo, Milano 2013; M. Guerrini, T. Possemato, *Linked data per biblioteche, archivi e musei: perché l'informazione sia del web e non solo nel web*, con un saggio di Carlo Bianchini e la consulenza di Rosa Maiello e Valdo Pasqui, prefazione di Roberto Delle Donne, Bibliografica, Milano 2015.

21. «Gli OCR attualmente in commercio si dimostrano molto efficienti in casi di riconoscimento di caratteri perfettamente conservati, ma per lettere sottoposte al logorio del tempo la loro percentuale di errore rimane talmente alta da renderli inutilizzabili. Nonostante siano state eseguite accurate operazioni di *training* per il riconoscimento di caratteri antichi, le condizioni di rumore, generalmente determinate da trasparenza dell'inchiostro dal verso al recto e viceversa, creano confusione nella catalogazione dei caratteri in tipi ben delineati. Il disturbo, infatti, genera una difficoltà nel ritrovamento dei contorni di ogni singola lettera da parte del sistema, che, pertanto, nelle fasi di *training*, chiede all'operatore di che gli venga fornito il valore ASCII di molti esempi dello stesso carattere, fatto che inevitabilmente causa incertezza nella classificazione di altri campioni che presentino anche leggere differenze grafiche rispetto ai tipi che sono stati utilizzati durante il *training set*.» Cfr. in rete il contributo di Andrea Bozzi <http://webilc.ilc.cnr.it/viewpage.php?sez=ricerca/id=77/vers=ita> (ultimo accesso 12 febbraio 2019).

22. I progetti, gli studi e le collaborazioni negli ultimi anni si sono moltiplicate. Per citarne un paio tra i centri/progetti più importanti: Istituto per la linguistica computazionale A. Zampolli del CNR di Pisa; l'AIDUC Associazione per l'informatica umanistica e la cultura digitale (l'ICCU è partner dell'AIDUC).

23. A. Ciula, F. Stella (a cura di), *Digital Philology and Medieval Texts*, Pacini, Ospedaleto 2007; C. Basile, M. Lana, *L'attribuzione di testi con metodi quantitativi: riconoscimento di testi gramsciani*, in "AIDAinformazioni", 1-2, 2008, pp. 165-83; P. Cotticelli Kurraz (a cura di), *Linguistica e filologia digitale: aspetti e progetti*, Edizioni dell'Orso, Alessandria 2011; A. Bernardelli, *Intertestualità*, La Nuova Italia, Firenze 2000; dello stesso autore si veda anche: *Che cos'è l'intertestualità*, Carocci, Roma 2013.

24. *Linked data* o dati collegati sono una modalità di pubblicazione di dati strutturati, collegati fra loro e quindi utilizzabili attraverso un tipo di interrogazione semantica. Si veda, tra gli altri: C. Bizer, T. Heath, T. Berners-Lee, *Linked Data: The Story So Far*, in "International Journal on Semantic Web and Information Systems", 5, 3, 2009, pp. 1-22.

25. *Big data* è l'insieme delle tecnologie e delle metodologie di analisi di dati massivi. Tra gli altri, Si veda: A. De Mauro, M. Greco, M. Grimaldi, *A Formal Definition of Big Data Based on Its Essential Features*, in "Library Review", 65, 3, 2016, pp. 122-35.

26. È chiaro che le figure professionali qui citate rappresentano la connessione di diversi saperi che non si acquisiscono solo con una laurea magistrale.

27. Naturalmente il panorama qui espresso non intende fornire un quadro esaustivo ma soltanto alcuni esempi di interazione direi... obbligata... tra varie tipologie di saperi, ugualmente utili alla valorizzazione, fruizione e promozione dei beni culturali materiali e immateriali. Aggiungerei inoltre il fattore determinante della conoscenza delle lingue straniere, di primaria importanza al fine di rendere internazionale un progetto di *digital curation*, filologia digitale ecc.

28. D. Fiornante, *Per una critica del testo digitale: Letteratura, filologia e rete*, Bulzoni, Roma 2018.

29. Siamo tra il 1986 e il 1987.

30. Si veda: <http://www.tei-c.org/About/history.xml> (ultimo accesso 12 febbraio 2019).

31. F. Meschini, *Edizioni critiche digitali: sul rapporto tra testo, edizione e tecnologia*, in "Digitalia", 24, 2014, pp. 24-42. Disponibile sul web all'indirizzo: <http://digitalia.sbn.it/article/viewFile/829/554> (ultimo accesso 12 febbraio 2019). Si veda anche il volume P. Italia, C. Bonsi (a cura di), *Edizioni Critiche Digitali; Digital Critical Editions: edizioni a confronto, comparing editions*, Sapienza Università Editrice, Roma 2016, consultabile al link http://www.editricesapienza.it/sites/default/files/5369_Italia_Bonsi_EdizioniCriticheDigitali.pdf (ultimo accesso 12 febbraio 2019).

32. Non esistono al momento sul mercato software che possano garantire una collazione completamente automatica. Juxta è raggiungibile al link: <http://www.juxtagoftware.org/> (ultimo accesso 12 febbraio 2019). Come sostengono Manuel Portela e António Rito Silva «The TEI format is not suitable to support the dynamic and concurrent change of its structure. Therefore, the implemented solution results from a compromise that integrates the traditional workflows of textual encoding for critical editing with the dynamics of Web 2.0 interactions». Cfr. degli autori citati: *Fernando Pessoa's Book of Disquiet as a Dynamic Digital Archive*, in Italia e Bonsi, *Edizioni Critiche Digitali; Digital Critical Editions: edizioni a confronto, comparing editions*, cit., pp. 32. Tra i software sviluppati in Italia, da ricordare, EVT (Edition Visualization Technology) nato all'Università di Pisa sulla base degli schemi pubblicati dal consorzio TEI. Lo spunto viene dal contesto del progetto Vercelli Book Digitale; il software permette anche a singoli studiosi, che non hanno la possibilità di affidarsi a un team di tecnici e informatici specializzati, di diffondere in rete i risultati del proprio lavoro filologico. Si vedano in tal senso EVT: <http://sourceforge.net/projects/evt-project/> e il progetto Digital Vercelli Book: <http://vbd.humn.et.unipi.it/> (ultimo accesso 12 febbraio 2019).

33. Si vedano: F. Ciotti, *Text Encoding as a Theoretical Language for Text Analysis*, in D. Fiornante, J. Usher (eds.), *New Media and the Humanities: Research and Applications. Proceedings of the First Seminar "Computers, Literature and Philology"*, Edinburgh, 7-9 September 1998, OUCS, University of Oxford, Oxford 2001; P. Robinson, *Electronic Editions for Everyone*, in W. McCarty (ed.), *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Open Book Publishers, Cambridge 2010, pp. 145-63.

34. Si veda: <http://www.juxtagoftware.org/> (ultimo accesso 12 febbraio 2019).

35. Si veda, tra gli altri: A. Bozzi, *Edizione elettronica dei testi e filologia computazionale*, in A. Stussi (a cura di), *Fondamenti di critica testuale*, il Mulino, Bologna 2006, pp. 207-32.

36. Il Nuovo Soggettario è lo strumento impiegabile nell'indicizzazione per soggetto di risorse di varia natura, realizzato a cura del Servizio Bibliotecario Nazionale di Firenze. È aderente ai principi stabiliti dall'International Federation of Library Associations and Institutions (IFLA) e alle indicazioni degli standard internazionali. Lo strumento è rivolto a biblioteche italiane (generali, specializzate, specialistiche) e, in particolare, a quelle che operano nell'ambito del Biblioteca Nazionale Centrale (SBN), così come a musei, mediateche, archivi, centri di documentazione. Il sistema Nuovo Soggettario è in continua evoluzione e accrescimento. La Bibliografia Nazionale Italiana (BNI) lo impiega dal 2007 (si veda: <http://thes.bnfc.firenze.sbn.it/> (ultimo accesso 12 febbraio 2019)).

37. Si veda: Commissione permanente per la revisione delle regole italiane di catalogazione (a cura di), *Regole italiane di catalogazione: REICAT*, ICCU, Roma 2009. Il PDF è disponibile al seguente indirizzo: <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/REICAT-giugno2009.pdf> (ultimo accesso 12 febbraio 2019).

38. E. Pierazzo, *Le edizioni digitali e l'analisi linguistica: i casi di Jane Austen e di Anton Francesco Doni*, in Italia e Bonsi, *Edizioni Critiche Digitali, Digital Critical Editions: edizioni a confronto, comparing editions*, cit., p. 20.

Le intestazioni sono state ricavate dai rispettivi OPAC nazionali. La ricerca è partita dal metopac KVK <https://kvk.biblio.thek.kit.edu> (ultimo accesso 12 febbraio 2019). Si è poi tenuto conto di VIAF e del Thesaurus del CERL. Chiaramente si tratta di una esemplificazione (il cantiere è attivo) e le intestazioni sono suscettibili di aggiornamenti. Per le intestazioni in altre lingue, si è tenuto conto degli OPAC delle rispettive Biblioteche Nazionali.

39. Tomasin, *L'impronta digitale*, cit., p. 137.