



Un secolo di scritture fuori dal canone. La bibliografia come modello informatico per la ricerca

Alessia Scacchi

Università di Roma 'La Sapienza'
alessia.scacchi@uniroma1.it

Abstract. The aim of this research is to reconstruct the Italian bibliography of texts not included in the literary canon of the 20th century. The sources surveyed for this purpose include SBN, InternetCulturale, AIB, resources belonging to international research centres, and Google Books, consulted to double-check questionable information. Building the online bibliography required the definition the structure of bibliographic data. The model adopted is based on the authority records issued by SBN, which associates authors with hyperlinks to the sources used to disambiguate the entity and validate the sources themselves. The survey includes first and subsequent editions of the texts. The objective of the research is to re-think the bio-bibliography of works belonging to the 20th century in the light of the Digital Humanities.

Abstract. La Ricerca, finalizzata alla costruzione di una bibliografia italiana delle scritture fuori dal canone del Novecento, consta di un sistema di tabelle che ricostruiscono un quadro di presenze atipiche nel panorama letterario del Novecento. Le fonti utilizzate per la ricerca sono molteplici: SBN, InternetCulturale, AIB, risorse di Istituti di Ricerca internazionali, oltre a Google-books per rintracciare informazioni o disambiguare dati controversi. La costruzione della bibliografia in formato elettronico ha dunque condotto a ripensare la strutturazione dei dati bibliografici al fine di ottenere il massimo grado di leggibilità e di chiarezza della cospicua mole di dati catalogata, ma anche la massima permeabilità ed adattamento dei dati grezzi immessi nel database. Seguendo lo schema della "scheda di autorità" proposto da SBN, per ogni autore/autrice il data base bibliografico contiene di prassi la dichiarazione delle fonti utilizzate sotto forma di link attivi, strumento essenziale per validare l'attendibilità delle fonti. La ricostruzione della bibliografia delle opere ha implicato inoltre la selezione non solamente della prima attestazione a stampa di ogni opera, ma delle nuove edizioni successive. Con tale ricerca è stato ripensato il concetto stesso di bio-bibliografia delle opere del Novecento, aggiornandone scopi e architettura, riflettendo infine sulle specificità introdotte dalle Digital Humanities.

Keywords: Literature, Representation, Resources and Interoperability, Text Analysis and Digital Objects Processing

1 Introduzione

Le parole contenute in una bibliografia rischiano di risultare ideologicamente disorientate, come sosteneva Saramago.¹ ma esse lo sono ancor di più rispetto a quelle contenute in ogni singolo testo letterario. Quindi, nel costruire e rendere fruibile un insieme di dati bibliografici è auspicabile determinare, come già proposto da Lotman in altro contesto, la grandezza della sua entropia e mirare a recuperare la sua complessa costruzione extratestuale.² Essa non si oppone in modo binario al testo artistico in sé, come neppure al parlante/scrivente di cui si vuole ricostruire la vicenda letteraria e intellettuale. Ciò è tanto più vero quando la prospettiva analitica prescelta per la ricerca bibliografica è rappresentata da un oggetto alquanto trasversale, ovvero la letteratura firmata da donne.

In questo caso nessuno strumento, fino ad oggi, ha unito il rigore scientifico alla ideazione e realizzazione di una bibliografia delle scritture sistematicamente estromesse dal canone del Novecento letterario italiano. Perciò la ricerca scientifica deve necessariamente porsi alcune questioni di metodo connesse a un uso consapevole delle fonti in merito alla loro verificabilità e attendibilità. Nel costruire una bibliografia che si avvalga degli strumenti informatici sono stati messi in discussione i primi strumenti di archiviazione messi a punto negli anni Novanta, due sistemi di tabelle diverse che ricostruivano un quadro di presenze atipiche nel panorama letterario dell'epoca senza interrogarsi sulla valenza culturale dello strumento digitale prescelto.

Fin dagli iniziali momenti dedicati all'organizzazione del lavoro è emerso un patrimonio cospicuo di pubblicazioni di scrittrici italiane novecentesche; patrimonio individuato già dalle prime indagini sistematiche condotte negli anni Novanta, che ammonta a centinaia di autrici e di opere. Come noto, le scrittrici vennero escluse dal canone delle prime antologie e storie della letteratura italiana post-risorgimentale,³ rafforzando così una concezione tutta maschile del fatto letterario. Nonostante ciò si ponevano come fattore trainante della ricostruzione critica e storiografica del secolo successivo. Ciò avveniva in anni di grande fermento culturale, anni in cui il panorama dell'editoria nascente della nuova Italia si andava delineando in maniera netta e inequivocabile. Molte infatti erano le scrittrici che pubblicavano opere d'ingegno di vario tipo, dai romanzi alle raccolte di novelle, dalle opere poetiche alla letteratura per l'infanzia, dalle antologie per le scuole alle letture critiche. Tutto ciò scomparve per dare spazio alle prove letterarie maggiormente commentate oltre che più studiate.

Oggi la critica ha l'onere, ma anche l'onore, di ricostruire una mappa di opere di varia natura, peraltro scarsamente definibili in quanto al genere letterario. Da questa esigenza nasce il presente lavoro, che ha come obiettivo proprio la riscrittura e il sostanziale ripensamento metodologico, ai fini di una più ampia e inclusiva ricostruzione, di un patrimonio dimenticato, che invece ha formato generazioni di studenti e di lettori primo-novecenteschi.

¹ Cfr. [4].

² Cfr. [2].

³ De Sanctis, seguito da poi Croce, furono i primi artefici nella costruzione del canone che emerge nelle *Storie della Letteratura italiana*, come del resto risulta in maniera chiara dagli scritti di Marina Zancan a partire dal volume *Il doppio itinerario della scrittura*, Torino, Einaudi, 1998.

2 Bibliografia delle Autrici del Novecento: fonti e struttura

BibliAN, il nome attribuito alla *Bibliografia delle Autrici del Novecento*, è frutto di un lavoro ventennale che nasce nell'ambito del progetto di ricerca *Scrittrici e intellettuali del Novecento* coordinato da Marina Zancan.⁴ Tra tutti i progetti dedicati alle scrittrici l'esperienza del database IWW (Italian Women Writers) è stato certamente un punto di riferimento anche se attualmente non è più aggiornato⁵.

Lo scavo nella documentazione bibliografica su materiale che non è sistematizzato e che non presuppone la differenza di genere come valore distintivo e definitorio, si è configurato, nei fatti, come un lavoro ricostruttivo; un restauro che si giova di nuovi materiali recuperando ciò che di vecchio e dimenticato si aveva negli archivi nazionali e internazionali. Si è dato vita, perciò, a un'archeologia del documento librario che riscrive la mappa dell'editoria dall'Unità ad oggi, ridefinendo campi d'interesse e di forza disegnati non dalla differenza di genere come pregiudizio, quanto piuttosto dalla molteplicità dell'offerta editoriale del secolo breve, nella sua densità di significati.

Del resto l'attuale panorama di progetti che si avvalgono di strumenti informatici che si occupano di scrittrici mette a disposizione numerose risorse; eppure, nel caso di *BibliAN*, la scelta è stata volutamente attenta e complessa e ha condotto a una serie di risultati che derivano proprio dalle questioni teorico-metodologiche, di cui si renderà conto nei paragrafi successivi.

Il lavoro di analisi e sistematizzazione della *Bibliografia delle Autrici*, è stato molto complesso visto che si avvale di quattromila record bibliografici relativi a opere letterarie di vario genere. Oltre a ciò, è elevato anche il numero di autrici attive tra il 1882 e il 2014. Il modello di bibliografia ha preso quindi in esame non solo nomi e date ma anche le numerose case editrici coinvolte nella promozione e diffusione di opere di donne dalla fine dell'Ottocento al nuovo millennio. Si vedano come esempio alcuni degli elementi presi in considerazione nella schedatura: la data di elaborazione del testo, che è stato possibile evincere in gran parte dei casi solamente grazie a uno studio approfondito sulla biografia e la bibliografia di ogni singola autrice inserita nel DB; oppure il campo dedicato ai premi letterari conferiti alla singola opera; così come la possibilità di registrare non solo il dato storico legato allo pseudonimo usato dall'autrice per pubblicare - molto noti sono alcuni esempi di scrittrici che utilizzano nomi maschili per avere maggiore libertà nella pubblicazione -⁶ ma anche il patronimico, così come il cognome del marito conservato come proprio sul frontespizio dei romanzi o dei racconti editi.

È possibile affermare che la prassi informatica aumenti le possibilità significative connesse con l'interpretazione poiché procede trasversalmente rispetto al *continuum* di qualsivoglia lista di record bibliografici. Ciò consente di attraversare il database secondo diverse linee di ricerca, producendo dati scientificamente molto rilevanti, se si pensa che le scrittrici

⁴ Progetto finanziato a più riprese dal 1999 dell'Università di Roma 'La Sapienza'. Cfr. <http://scrittrici900.uniroma1.it>.

⁵ Il Progetto dell'Università di Chicago, non più aggiornato dal 2010, non sembra avere un piano a lungo termine di mantenimento e sia il motore di ricerca che l'intera infrastruttura appaiono obsolete. <https://www.lib.uchicago.edu/efits/IWW/> [7].

⁶ Si veda come esempio molto noto il caso di Beatrice Speraz che prende il nome di Bruno Sperani per non subire critiche dal pubblico e dalla società dell'epoca.

novacentesche riescono ad accedere alle redazioni di poche case editrici e, di alcune, aspirano al massimo a conquistare solamente determinate collane, quasi ghezzandosi entro tali confini, fors'anche comodi.

Questi dati, assenti da altre risorse diffuse in rete, consentono di ricostruire un quadro storiografico più complesso di quanto si creda e di ridisegnare la mappa delle scritture del Novecento proprio grazie a nuovi modelli e strumenti interpretativi.

Per ricostruire scientificamente il doppio itinerario della scrittura novecentesca indicato da Zancan già nel 1998,⁷ la raccolta e la riorganizzazione delle informazioni in formato digitale si delinea come una rete di senso che oltrepassa le differenze di genere. L'originaria forma tabulare assunta dal materiale grigio da cui parte il presente lavoro, quindi, si trasforma in una infrastruttura di significato quando viene riorganizzata. Il database che ne è il frutto, *BibliAN* appunto, veicola flussi di informazioni, di spiegazioni e di ipotesi.⁸

La bibliografia ottenuta è impostata come una base di dati che intende esplorare le scritture di donne dal 1881 al 2016 - anno in cui è formalmente terminata la ricerca -⁹ garantendo nel tempo un ambiente che possa incidere significativamente e indefinitamente sui processi di formazione e gestione delle memorie.

Per una reinterpretazione dei dati che sia attuale e superi l'«ennesimo trionfo del tipografo sull'autore» è stato però necessario riflettere sulle metodologie, incentivare un lavoro d'équipe e condividere le informazioni raccolte e la discussione metodologica.

Volendo tracciare il percorso di ricerca, si è stabilito di suddividere il lavoro utilizzando due categorie principali: autrici e opere. Per ogni autrice, quindi, è stato seguito lo schema della "scheda di autorità" proposto da SBN in modo da valorizzare e al contempo mappare la biografia di ogni romanziera o poetessa presente nel database bibliografico.¹⁰ Perciò ogni record contiene la dichiarazione delle fonti utilizzate sotto forma di link attivi, strumento essenziale per validare l'attendibilità delle fonti. La ricostruzione della bibliografia delle opere ha implicato la selezione non solamente della prima attestazione a stampa di ogni opera ma anche delle nuove edizioni successive¹¹. Dalle schede SBN vengono tratte informazioni biografiche fondamentali, come le date di nascita e morte, le indicazioni relative al patronimico e allo/agli pseudonimi, l'attività intellettuale svolta accanto alla scrittura di prosa e poesia; sono presenti nel database, infatti, non solamente scrittrici, ma giornaliste, saggiste, maestre, docenti universitarie, che segnano il Novecento della letteratura italiana pur rimanendo talvolta sconosciute ai più.

3 Le metodologie e le tecnologie

Se le fonti dei contenuti sono utili alla valutazione dei risultati di un progetto bibliografico digitale volto alla ricostruzione delle presenze letterarie femminili nel 900 italiano, tanto più lo sono le scelte che hanno guidato lo sviluppo tecnologico della base di dati.

⁷ Cfr. [6].

⁸ Cfr. [5].

⁹ Seguendo la periodizzazione di Zancan.

¹⁰ La ricostruzione delle biografie delle letterate del Novecento rappresenta un primo passo per la restituzione di un profilo bio-bibliografico delle autrici.

¹¹ Datando la fine del progetto al 2016.

Il progetto ha preso avvio con il lavoro di sistematizzazione delle informazioni bibliografiche registrate inizialmente come tabelle di fogli di calcolo Google Drive, condivisi da tutti i partecipanti alla ricerca. Per trasformare questi dati in una banca dati a tutti gli effetti si è preferito adottare la tecnologia XML piuttosto che un classico database relazionale. In particolare, il linguaggio di marcatura utilizzato è stato quello XML TEI (*Text Encoding Initiative*) nella versione P5, una scelta in certo senso obbligata, visto che si tratta di uno standard di codifica riconosciuto nella comunità scientifica internazionale delle *Digital Humanities*.¹²

La versione XML/TEI dei dati bibliografici viene generata in modo automatico importando i dati in una tabella di Excel, e successivamente utilizzando lo strumento di esportazioni in formato XML del programma Microsoft, che richiede la creazione di un XML Schema per modellare la struttura dei dati presi in considerazione. Per l'archiviazione ed elaborazione del file XML risultante è stato utilizzato uno dei più avanzati ed efficienti database XML nativi, BaseX.¹³ Tale strumento si presenta come un sistema di archiviazione, interrogazione e visualizzazione di grandi documenti e collezioni codificate in XML ed è pensato per essere consultabile anche tramite Web.¹⁴ BaseX ha inoltre l'indubbia qualità di essere un software multiplatforma, oltre a essere un prodotto *open-source*: due peculiarità che lo rendono ideale in ambito accademico, viste le finalità di ricerca.

Volendo specificare e illustrare le caratteristiche di BaseX è importante far notare che esso si presenta come un motore di database XML, al contempo robusto nelle sue specifiche tecniche ed efficiente quanto a velocità e capacità di gestione dei dati. L'interrogazione del database si basa su una affidabile implementazione dello standard W3C XQuery,¹⁵ ed è facilitata dalla disponibilità di un editor che in tempo reale evidenzia la sintassi e fornisce un feedback sugli eventuali errori di sintassi. Questo accelera lo sviluppo e il miglioramento delle applicazioni XQuery/XML, mentre l'editor delle interrogazioni aiuta l'utente nella valutazione della sintassi in tempo reale.

BaseX adotta una architettura *client/server* per gestire le operazioni simultanee di più fruitori e il *front-end* si caratterizza per una serie di visualizzazioni gerarchiche che consentono di esplorare i dati. Per l'accesso singolo al database è utilizzabile un servizio *WebDAV* che consente di archiviare, modificare e organizzare rapidamente le risorse con un semplice file manager abilitato.

Questo strumento consente di realizzare delle *query* sul documento proprio come avviene nel caso dei tradizionali database relazionali. In questo caso l'utilizzo del linguaggio XQuery ha reso il recupero delle informazioni dal documento XML molto agevole. Essendo costituito da una sintassi semplice e facilmente leggibile si sono formulate delle query sui dati facendo emergere delle peculiarità attraverso delle interrogazioni incrociate direttamente sulle annotazioni, i marcatori, della bibliografia.

¹² Per la TEI si veda <http://www.tei-c.org>.

¹³ <http://basex.org/>.

¹⁴ Per il momento la bibliografia non è ancora disponibile al pubblico. L'accesso sarà messo a disposizione degli utenti tramite una interfaccia Web inserita nell'ambito del sito web del Laboratorio CRILeT [8].

¹⁵ XQuery 3.1 per cui è previsto un supporto per l'ultimo aggiornamento W3C ufficiale e le raccomandazioni per la modalità Full Text; conforme alle raccomandazioni W3C XPath, compresi gli aggiornamenti.

Per quanto riguarda le visualizzazioni, che sono interattive, esse riescono a supportare documenti XML di grandi dimensioni, cosa evidentemente molto utile in fase di elaborazione dati; sono infatti disponibili un numero notevole di interfacce di cui si sta vagliando l'utilità specifica ai fini della visualizzazione da parte dell'utente finale.¹⁶

Le soluzioni per la ricerca *full text* rendono BaseX uno strumento ideale per recuperare le informazioni testuali presenti nella base di dati: si ha la possibilità di utilizzare caratteri jolly, riconoscere maiuscole e minuscole, segni diacritici, punteggio TF/IDF e segnalare una stop-list. Gli indici di testo, attributo, *full-text* e *Path-Summary* si sono rivelati piuttosto compatti, accelerando così il processo di valutazione dei risultati secondo gli ordini di grandezza prescelti. In aggiunta a ciò, ogni procedura di valutazione è visualizzabile con singoli passaggi e conseguenti accessi agli indici, mentre i piani di esecuzione sono serializzabili in testo semplice, in XML e nel formato Word.

Infine è molto sicura la gestione degli utenti finali del database. Infatti le autorizzazioni globali e locali possono essere assegnate agli utenti che si collegano al server utilizzando un'autenticazione sicura, garantita peraltro dalla registrazione di tutte le transazioni, ovvero le attività di lettura, scrittura e server vengono registrate in ordine cronologico.

Veniamo dunque alla visualizzazione dei risultati che, negli intenti, si voleva agevole per l'utente medio, ma anche interessante e produttiva per un utente esperto e abituato a strumenti di ricerca digitali. La pagina si suddivide in riquadri (*frame*), nei quali viene offerto il risultato della query in differenti modalità visive: a partire dall'albero informativo per giungere al grafico distributivo. Ogni frame può essere impostato e implementato dall'utente a seconda della specifica indagine da cui si vuole trarre materiale utile alla ricerca scientifica. Ciò cambia radicalmente l'approccio del fruitore poiché i dati possono essere mescolati e fatti interagire a seconda del background culturale e scientifico dello/a studioso/a.

Dunque, ciò che in database come IWW erano tre distinte componenti - un indice autore, un indice delle edizioni e un indice full text - all'interno di *BibliAN*, si tramuta in una risorsa aperta. Per questo si è scelto di non presentare i testi delle autrici in formato digitale, come fa invece IWW, poiché essi, codificati in HTML come avviene nel sito dell'Università di Chicago, sono materiale statico, risultano difficilmente fruibili dall'utente medio, oltre a non essere analizzabili da alcun software di *text analysis* o *information retrieval*.

In *BibliAN*, al contrario, la mole di dati è in continua evoluzione e si sta lavorando alla creazione di un gruppo di studio che si concentri sulle singole figure intellettuali che via via si vanno rintracciando nel panorama letterario novecentesco, al fine di offrire agli studiosi e alle studiose una prospettiva più ampia e più completa sul secolo. Infine pare doveroso, dopo vent'anni di studi di Informatica Umanistica, proporre un sistema costantemente aggiornato che non segni solamente il *limes* delle terre conosciute, quanto piuttosto si proietti verso un più ignoto ma certamente più interessante universo da scoprire.

¹⁶ Sono disponibili le seguenti interfacce REST/RESTXQ, WebDAV, XQJ, Xml:DB.

4 Alcuni risultati: la fine che è un inizio

I risultati attesi sono stati di vario tipo, in quanto vari erano i fronti su cui si muoveva la ricerca, ormai in via di completamento.¹⁷ Questi sono: la Storia della Letteratura italiana, la Critica letteraria, la Teoria della letteratura e, non ultima (nonostante non sia ufficialmente riconosciuta come disciplina), l'Informatica umanistica.

Un primo riscontro si è avuto trasversalmente rispetto agli ambiti disciplinari poiché la bibliografia in formato elettronico ha valorizzato questo strumento come metodo di apprendimento e rielaborazione dei contenuti specifico ed unico, come mezzo di comprensione e di conservazione a lungo termine del patrimonio conoscitivo in ambito umanistico, come risorsa per ripensare e mettere in discussione il Canone in maniera completamente nuova. Inoltre, un simile prodotto può essere utilizzato nelle scuole superiori per fornire una prospettiva 'altra' sul Novecento, offrire uno sguardo critico sul materiale narrativo tradizionale tramite l'ampliamento della proposta letteraria.

Secondo passaggio obbligato è il riconoscimento di un immenso lavoro di ricerca a partire dallo scavo degli archivi d'autore, fino ai servizi di OPAC e ai database già esistenti in rete ma, come accennato, datati per quanto riguarda l'architettura complessiva e la visione prospettica del progetto. Il formato digitale dei dati di *BibliAN* si configura esso stesso come una specifica metodologia di ricerca poiché è stato allestito *ad hoc* uno schema bibliografico da utilizzare sia per le indicizzazioni di opere, sia per la biografia delle autrici che può essere utilizzato, ovviamente, senza alcuna distinzione di genere.

Infine, è estremamente stimolante per gli studiosi e le studiose che hanno lavorato e lavorano al progetto, la possibilità di mettere a frutto un lavoro tutto italiano che consenta alla comunità scientifica internazionale di beneficiare dei risultati comprovati da un approccio scientifico di lunga data. La speranza è che il lavoro avviato da questa ricerca faccia emergere la bibliografia ottenuta come un fenomeno avente significato in sé. Nel disordine della rete è dunque possibile evidenziare e far emergere coerenza e unitarietà, oltre che scientificità, della modellizzazione informatica del dato testuale e tabulare.

La costruzione della bibliografia in formato digitale conduce a ripensare, infatti, la strutturazione del dato bibliografico, al fine di ottenere il massimo grado di leggibilità e di chiarezza della cospicua mole catalografica, ma anche di garantire la massima permeabilità ed adattamento dei dati grezzi immessi nel database.

Potremmo dunque sostenere che in tal modo si sia avviata una trasposizione di contenuti dal formato tabulare a quello della base di dati in formato elettronico. Saltano i confini delle singole celle, esplodono le possibilità di significazione ma, al contempo, si rivelano pregni di senso tutte le modificazioni e la possibilità stessa di rivedere ogni osservazione empirica veicolata dal database. Come osservazione finale, il modo di procedere che abbiamo adottato in *BibliAN* ci fa pensare alla prima regola metodologica proposta da Popper: «Il giuoco della scienza è, in linea di principio, senza fine. Chi, un bel giorno, decide che le asserzioni scientifiche non hanno più bisogno di nessun controllo, e si possono ritenere verificate definitivamente, si ritira dal giuoco».¹⁸

¹⁷ Sono in atto le procedure per ottenere i permessi per l'utilizzo di un dominio Sapienza, all'interno del sito della ricerca di Marina Zancan "Scrittrici e intellettuali del Novecento" [9] e del Laboratorio CRILeT [8].

¹⁸ Cfr. [3].

References

1. Bauman, Zygmunt. *Modernità liquida*. Bari: Laterza, 2000.
2. Lotman, Jurij, M. *La struttura del testo poetico*. Milano: Mursia, 1972.
3. Popper, Karl, R.. *Logica della scoperta scientifica*. Torino: Einaudi, 1970.
4. Saramago, José. *Storia dell'assedio di Lisbona*. Torino: Einaudi, 2000.
5. Weinberger, David. *Elogio del disordine. Le regole del nuovo mondo digitale*, Milano, BUR, 2010.
6. Zancan, Marina, *Il doppio itinerario della scrittura*. Torino: Einaudi, 1998.
7. "Italian Women Writers", <https://www.lib.uchicago.edu/efts/IWW/>.
8. "Laboratorio CRILET", <http://crilet.uniroma1.it>.
9. "Scrittrici e intellettuali del Novecento". <http://scrittrici900.uniroma1.it/>.

Ultima consultazione URLs: 22/03/2018