

Dig *Italia*

Anno XIX, Numero 1 - **2024**

Rivista del digitale nei beni culturali

ICCU-ROMA



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
www.iccu.sbn.it

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista
è ammessa con obbligo di citazione

«*Digitalia*», rivista del digitale nei beni culturali, è una rivista *peer reviewed*
e segue il codice etico delle pubblicazioni

I contributi possono essere proposti alla rivista dagli autori o su iniziativa del Comitato Scientifico. Gli articoli vengono sottoposti al Comitato di Redazione per un primo accertamento sulla corrispondenza con i campi di ricerca della rivista. I contributi delle sezioni Saggi e Progetti vengono indirizzati in forma anonima ad almeno uno studioso di comprovata competenza sui temi affrontati. I revisori fanno pervenire i loro giudizi alla redazione (favorevole alla pubblicazione, favorevole con modifiche/miglioramenti, non favorevole). Se il giudizio finale è positivo, viene comunicata agli autori l'accettazione del contributo, insieme ad eventuali indicazioni suggerite dai valutatori, di cui si garantisce comunque l'anonimato.

Digitalia

Rivista del digitale nei beni culturali
ISSN 1972-621X
Anno XIX, Numero 1 - Giugno 2024

In copertina:

L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.C. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicinabile all'Apollo di Leochares (IV sec. a.C.)

Direttore Fondatore

Marco Paoli

Direttore Responsabile

Simonetta Buttò

Comitato di Redazione

Amalia Maria Amendola
Valentina Atturo
Laura Borsi
Flavia Bruni
Elisabetta Caldelli
Elisabetta Castro
Silvana de Capua
Vilma Gidaro
Maria Cristina Mataloni
Lucia Negrini
Federica Olivotto
Angelo Restaino
Elisa Sciotti
Alice Semboloni
Vittoria Tola
Maria Lucia Violo

Grafica & Impaginazione

MLA&Partner - Roberta Micchi

Produzione e Stampa

Istituto Poligrafico e Zecca dello Stato S.p.A.
Roma

Editore

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane
e per le informazioni bibliografiche
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49.210.425
email: ic-cu.digitalia@cultura.gov.it
<https://digitalia.cultura.gov.it>



Comitato Scientifico

Stefano Allegrezza
Osvaldo Avallone
Giovanni Bergamin
Dimitri Brunetti
Simonetta Buttò
Rosaria Campioni
Laura Ciancio
Gianfranco Crupi
Andrea De Pasquale
Pierluigi Feliciati
Marina Giannetto
Maria Guercio
Mauro Guerrini
Klaus Kempf
Maurizio Messina
Maria Cristina Misiti

Laura Moro
Maria Teresa Natale
Marco Paoli
Don Valerio Pennasso
Massimo Pistacchi
Marco Pizzo
Paola Puglisi
Roberto Raieli
Gino Roncaglia
Maria Letizia Sebastiani
Giovanni Solimine
Laura Tallandini
Anna Maria Tammaro
Costantino Thanos
Antonella Trombone
Paul Gabriele Weston

SOMMARIO

giugno 2024

SAGGI

- Pubblica amministrazione e coscienza ambientale:
la digitalizzazione tra ecosostenibilità e falsi miti**
di Flavia Bruni 9
- I requisiti per l'addestramento degli strumenti
di Intelligenza Artificiale e il deposito legale
delle risorse digitali: una riflessione sul contesto
normativo italiano e il ruolo delle biblioteche**
di Chiara Storti 23
- Il riversamento di formato elettronico
tra standard internazionali e Linee guida
sulla formazione, gestione e conservazione
di documenti informatici dell'AgID**
di Stefano Allegrezza 36
- La trascrizione automatica di documenti a stampa
antichi. Appunti per un modello di riconoscimento
della tipografia in corsivo**
di Stefano Bazzaco 63

PROGETTI

- I libri Sapienza parlano. Un progetto per l'accessibilità
dei libri della Sapienza alle persone
con disabilità visive e DSA**
di Laura Armiero, Gianfranco Crupi, Angela Di Iorio,
Loredana Di Lucchio, Agnese Galeffi, Alessandra Gulotta,
Roberto Raieli, Valentina Rovacchi 89
- Quel che resta di un codice: nuovi orizzonti
per la catalogazione dei frammenti in MOL 2.0**
di Elisabetta Caldelli, Lucia Negrini 110
- IVS Comune online: proposta di un sistema
integrato per lo studio della trasmissione
dei testi (del diritto)**
di Alessandra Panzanelli 124
- Una base dati per lo studio del sistema
del privilegio librario nell'Europa moderna**
di Marius Buning, Alberto José Campillo Pardo,
Alica-Nana Citron, Andrea Ottone 147

**Tecnologie GIS nell'ambito
dei beni culturali ecclesiastici** 159
di Claudio Camuto, Laura Gavazzi,
Vittoria Rebuttini, Stefania Ricci

**La prima campagna di digitalizzazione
degli archivi del Venerabile Collegio Inglese di Roma** 171
di Renaud Milazzo

DOCUMENTI E DISCUSSIONI

**Computabilità e traduzione interlinguistica:
considerazioni sull'impiego dell'Intelligenza
Artificiale in ambito traduttivo** 183
di Marika Angelucci

**Tra bellezza e pericolo: le trappole
delle intelligenze artificiali nell'arte** 198
di Gianluca Garofalo

SEGNALAZIONI

**Il progetto ADELE e il tool
per l'autovalutazione digitale** 209
di Flavia Massara

**Un ponte tra scienza e società:
i nuovi prodotti editoriali digitali
del Dipartimento scienze umane e sociali,
patrimonio culturale del CNR** 211
di Giulia Antonini

**Il Geoportale Nazionale per l'Archeologia.
Presentato il 26 gennaio 2024 un bilancio
del progetto a sei mesi dalla pubblicazione** 215
di Valeria Boi

Saggi

Pubblica amministrazione e coscienza ambientale: la digitalizzazione tra eco-sostenibilità e falsi miti

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00091

Flavia Bruni

Università "G. D'Annunzio" – Chieti-Pescara

Partendo dall'analisi semantica del termine "dematerializzazione", elemento chiave del Codice dell'amministrazione digitale (CAD) entrato in vigore il 1° gennaio 2006, e di altre testimonianze, l'articolo vuole richiamare l'attenzione sull'impatto ambientale delle tecnologie digitali, problema rilevante anche per le professioni del patrimonio culturale ma ancora trascurato nel recente Piano nazionale di digitalizzazione del patrimonio culturale (PND).

«You would expect digital information replacing print to result in a reduced carbon impact. That may not be the case due to the so called rebound effect – the idea that when something (in this case, data) becomes cheaper and more carbon-efficient to do, we end up simply doing more of it».

Mike Berners-Lee,
How bad are bananas?
The carbon footprint of everything

1. Premesse

A ll'inizio del terzo millennio, molte questioni in territorio digitale rimangono ancora irrisolte. La cronica mancanza di coordinamento generale, particolarmente evidente nel campo delle *digital humanities*, ha spesso segnato il destino di progetti arenatisi su ostacoli già risultati fatali ad analoghe imprese precedenti; a minare la credibilità e la sussistenza dei progetti digitali è, sempre più spesso, la mancanza di sostenibilità, concetto che coinvolge molteplici piani interconnessi: da quello meramente tecnologico della disponibilità delle piattaforme e infrastrutture a quello delle decisioni strategiche, sullo sfondo di quello, fondamentale e determinante, del rischio del tracollo ambientale¹. Sebbene abbia esordito nel di-

¹ William Kilbride, *Memories from the Anthropocene: digital preservation in a time of climate crisis*, Digital Preservation Coalition, 4 November 2022: <<https://www.dpconline.org/blog/memories-from-the-anthropocene-digital-preservation-in-a-time-of-climate-crisis>>.

battuto scientifico piuttosto di recente, la questione dell'inquinamento delle tecnologie informatiche si impone con urgenza, visto il rischio concreto di pregiudicare l'esistenza stessa del pianeta. La crisi in atto si manifesta sempre più frequentemente e in varie forme, richiedendo una radicale revisione dell'agire quotidiano in quanto individui e in quanto società². Un'accresciuta sensibilità nei confronti delle tematiche ambientali è all'origine della novella degli articoli 9 e 41 della nostra Carta costituzionale avvenuta l'8 febbraio 2022, che ha introdotto tra i principi fondamentali la tutela dell'ambiente, della biodiversità e degli animali³. All'interno della sconfinata letteratura sull'impatto rivoluzionario delle tecnologie digitali su molteplici ambiti della nostra vita, quello specifico dell'incidenza sull'emergenza energetica e climatica pare invece essere un argomento ancora poco battuto. Partendo dall'analisi semantica della documentazione relativa alla cosiddetta dematerializzazione in atto nella pubblica amministrazione, questo contributo propone alcuni spunti per una riflessione costruttiva a fronte di criticità che hanno destato finora uno scarso interesse⁴.

2. L'Idea digitale

Una conoscenza superficiale dei processi informatici, combinata con la mancanza di contatto diretto con i supporti magnetici delle piattaforme remote (*in cloud*), la cui reale distanza dall'utilizzatore è annullata, a livello percettivo, dalla velocità dei collegamenti in rete, alimenta l'assimilazione di tutto ciò che è digitale a un'entità eterea. Sebbene un dato digitale sia essenzialmente «un fenomeno puramente elettromagnetico, che viene reso visibile attraverso appositi software» all'interno di un ambiente unidimensionale⁵, non mi sembra del tutto corretto considerarlo

² Si veda l'Agenda 2030 per lo Sviluppo Sostenibile, in particolare gli obiettivi 12 (*Garantire modelli sostenibili di produzione e di consumo*) e 13 (*Adottare misure urgenti per combattere il cambiamento climatico e le sue conseguenze*):

<<https://www.agenziacoesione.gov.it/comunicazione/agenda-2030-per-lo-sviluppo-sostenibile/>>.

³ <https://www.riformeistituzionali.gov.it/it/la-legge-costituzionale-in-materia-di-tutela-dell-ambiente/>.

⁴ L'idea di riprendere e armonizzare riflessioni scaturite in momenti e contesti diversi è maturata dalla mia partecipazione all'Europeana Climate Action Community steering group, <<https://pro.europeana.eu/page/climate-action-community>>, e dalla relativa condivisione di obiettivi e materiali; ringrazio Nicola Barbuti, Simonetta Buttò, Simona Casciano, Paola Castellucci, Stella Di Fazio, Manuela Grillo, Tiziana Mancinelli, Paolo Monella, Giulia Perucchi e Federico Valacchi per l'incoraggiamento e le osservazioni ricevute sulle bozze di questo contributo anche contestualmente alla presentazione nell'ambito del XII convegno annuale AIUCD (Siena, 6 giugno 2023), di cui un abstract esteso è pubblicato negli atti del convegno: Flavia Bruni, *Dematerializzazione e sostenibilità: prospettive ideali e problemi concreti*, in: *La memoria digitale. Forme del testo e organizzazione della conoscenza*. Atti del XII convegno annuale AIUCD (Università di Siena, 5-7 giugno 2023), a cura di E. Carbé, G. Lo Piccolo, A. Valenti, F. Stella (AIUCD, 2023), url open access:

<https://amsacta.unibo.it/id/eprint/7721/1/atti_aiucd_2023.pdf>. p. 327-331.

⁵ Tito Orlandi, *Informatica testuale: teoria e prassi*, Roma-Bari: Laterza, 2010, p. 15-16.

qualcosa di completamente “privo di materialità”. Per quanto volatili⁶, i dati digitali sono possibili e fruibili grazie a processi di codifica e decodifica, di trasmissione e di conservazione che hanno bisogno di un supporto materiale.

Un’analisi delle problematicità della conservazione digitale è particolarmente rivelatrice delle implicazioni su molteplici livelli non solo, o non tanto, da un punto di vista pratico, quanto piuttosto da quello concettuale. A rendere complessa la questione è prima di tutto l’alterabilità dei supporti magnetici, che può comprometterne la leggibilità; a ciò si aggiunge l’obsolescenza dei formati, che, pur ponendosi nell’altro ambito delle scelte attuate in fase di codifica per la creazione, manipolazione e trasmissione dei dati, complica la predisposizione di misure idonee sui diversi fronti. La confusione dei due piani ha indotto Carlo Federici ad affermare che

«la conservazione si identifica con la salvaguardia dei *materiali* costituenti il bene culturale i quali supportano il testo e conferiscono ad esso quella forma che ne consente la percezione. In ogni caso ciò vige solo per gli originali – o al massimo per le riproduzioni “analogiche” – poiché *la riproduzione digitale, essendo priva di materialità, sfugge alle leggi della conservazione, tant’è che la sua trasmissione al futuro si basa sostanzialmente su procedure di copia*»⁷.

La visione di un conservatore/restauratore, abituato a prendersi cura della materialità degli oggetti e distinguere ontologicamente la sostanza di ciascun esemplare da qualsiasi sua riproduzione, potrebbe semmai portare a riflettere sull’uso del termine “originale” riferito al digitale; ma anche nel mondo di bit e byte non si può prescindere dalla conservazione materiale del supporto, che rimane indispensabile per la sussistenza dei dati che ospita e veicola pur nel continuo processo di aggiornamento, che è effettivamente necessario, ma per garantirne l’accessibilità nel corso del tempo.

Nell’ambito di ciò che si definisce conservazione del digitale (*digital preservation*) si distinguono in effetti più filoni, in base alle sfaccettature che la disciplina assume in relazione ai vari contesti pertinenti, mentre il concetto stesso di conservazione non è privo di ambiguità⁸. Già nel 2007 Maria Guercio rilevava la complessità di prendere in considerazione «i molteplici aspetti connessi alla conservazione digitale, tra cui ad esempio i formati dei file, i servizi di registrazione dei formati medesimi, i modelli organizzativi per la gestione di depositi digitali», sottolineando contestualmente come la letteratura esistente risultasse vasta ma frammentaria⁹.

⁶ Ivi, p. 59.

⁷ Carlo Federici, *La conservazione dei materiali librari*, in: *Biblioteconomia: principi e questioni*, a cura di G. Solimine, P. G. Weston, Roma: Carocci, 2007, p. 379-394: p. 383, corsivo mio.

⁸ Maria Guercio, *Conservare il digitale: modello nazionale e contesto internazionale*, «DigitCult – Scientific Journal on Digital Cultures», 1, 2 (July 2016), p. 19-26, <<https://digitcult.lim.di.unimi.it/index.php/dc/article/view/10>>.

⁹ Maria Guercio, *La conservazione delle memorie digitali*, in: *Biblioteconomia: principi e questioni*, a cura di G. Solimine, P. G. Weston, Roma: Carocci, 2007, p. 395-412: 395.

Nel quindicennio successivo, gli archivisti informatici hanno svolto molto lavoro di elaborazione teorica e pratica lungo le direttrici già delineate, focalizzandosi da una parte sulla necessità di garantire l'accessibilità e fruibilità a lungo termine delle risorse tramite l'adozione di standard e modelli in grado di prevenire l'obsolescenza di formati e applicazioni, dall'altra sull'elaborazione di procedure di trasmissione che assicurino l'autenticità e l'integrità dei documenti¹⁰. Ciò ha determinato una polarizzazione intorno a temi e problemi riferibili al patrimonio archivistico, mentre sono rimasti sostanzialmente inesplorati, sia in ambito accademico sia professionale, terreni affini ma distinti quali la conservazione delle risorse web, delle biblioteche digitali (la cui definizione stessa rimane spesso ambigua e controversa anche nella letteratura di settore) e della documentazione bibliografica¹¹. Manca ancora, a livello nazionale ma per lo più anche internazionale, una strategia chiara e condivisa su cosa conservare e in che modo: basti pensare ai progetti di ricerca, che sempre più frequentemente realizzano non solo pubblicazioni più o meno tradizionali ma anche prodotti fruibili sul web (banche dati, siti dinamici, mappe interattive, edizioni critiche digitali e via dicendo) che, nella maggior parte dei casi, sono eliminati dai server accademici al termine del rapporto dei responsabili con i rispettivi atenei. Nonostante una cognizione ormai diffusa di questo problema, solo in casi sporadici sono state messe in atto contromisure concrete all'oblio digitale: tra questi si segnala il progetto per la conservazione del sito web dell'Istituto centrale per gli archivi (ICAR) nella sua evoluzione tra 2008 e 2016¹² a integrazione di

¹⁰ M. Guercio, *La conservazione delle memorie digitali*, cit., p. 397-398; per una panoramica più completa sull'argomento cfr. Ead., *Conservare il digitale. Principi, metodi e procedure per la conservazione a lungo termine di documenti digitali*, Roma-Bari: Laterza, 2013.

¹¹ M. Guercio, *Conservare il digitale: modello nazionale e contesto internazionale*, cit., p. 20; Maria Teresa Biagetti, *Le biblioteche digitali: tipologie, funzionalità e modelli di sviluppo*, Milano: FrancoAngeli, 2019, p. 7-18; e Gilberto Marzano, *Conservare il digitale. Metodi, norme, tecnologie*, Milano: Bibliografica, 2011. Cfr. anche T. Orlandi, *Informatica testuale*, cit., p. VIII: «lo stesso problema della conservazione dei testi già digitali in ambiente digitale solo recentemente ha trovato la dovuta attenzione presso le discipline archivistiche, ed è lontano dall'aver trovato proposte convincenti di soluzione».

¹² <<https://icar.cultura.gov.it/attivita-e-progetti/progetti-icar-1/la-conservazione-del-sito-web-dellistituto-centrale-per-gli-archivi-2008-2016>>: «Il settore dei beni culturali ha prodotto e continua a produrre una quantità sempre maggiore di contenuti digitali che hanno bisogno di essere archiviati, conservati e tutelati nel tempo in modo affidabile per consentire che queste risorse possano essere recuperate e riutilizzate in maniera efficace e possano essere utilizzate per le future ricerche storiche. Le problematiche relative alla conservazione del patrimonio digitale non sono state prese in sufficiente considerazione nella stragrande maggioranza delle iniziative di digitalizzazione del patrimonio e in quelle di costruzione di contenuti culturali, per cui di frequente si assiste alla loro scomparsa o impossibilità d'uso dei siti web, con la conseguente perdita della loro valenza culturale e storica e dispendio di risorse umane ed economiche. Consapevole dell'importanza della conservazione dei siti web, quali serbatoi di contenuti culturali e testimonianza storica dell'attività e delle strategie di comunicazione dei soggetti che li hanno promossi, l'Istituto Centrale per gli Archivi, nell'occasione della realizzazione di un nuovo sito web, aggiornato nelle tecnologie e nei contenuti, ha ritenuto necessario affrontare il problema di conservare quello precedente, nella sua interezza e nelle sue varie componenti (pagine web, oggetti digitali, basi di dati), applicando i principi, le metodolo-

quanto realizzato nell'ambito di Internet Archive, nato proprio con l'intento di documentare il patrimonio digitale garantendo un accesso a risorse non più disponibili in rete, o a versioni precedenti di risorse tuttora online, seppure per forza di cose limitato ad alcune pagine in una versione spesso semplificata¹³.

Quella espressa da Federici rimane comunque, in modo più o meno esplicito, una concezione ancora largamente diffusa delle risorse digitali, siano esse native o derivate da oggetti analogici. Per qualcuno, la presunta assenza di materialità, insieme ad altri fattori, è sufficiente a giustificare una gerarchia ontologica tra le risorse digitali e quelle analogiche, ovvero tangibili¹⁴.

3. Dematerializzazione o smaterializzazione? Breve excursus ideologico-semantico

Su questi presupposti si innesta il concetto di "dematerializzazione", cardine della riforma in atto per migliorare l'efficienza della pubblica amministrazione dall'entrata in vigore, il 1° gennaio 2006, del Codice dell'amministrazione digitale, noto come CAD (d.lgs. 82/2005, successivamente modificato e integrato dal d.lgs. 179/2016 e poi dal d.lgs. 217/2017)¹⁵. La dematerializzazione faceva la sua comparsa come elemento caratterizzante il neonato CAD:

gie e le tecniche del web archiving, così come si è andato definendo nel corso dell'ultimo quindicennio. Il progetto, ideato e realizzato nel dicembre 2016 da Costantino Landino, collaboratore dell'Istituto, si è articolato in due fasi principali: la prima dedicata all'harvesting e alla conservazione del sito web e la seconda alla predisposizione dei servizi di visualizzazione del sito archiviato».

¹³ <<https://archive.org/>>; su Internet Archive si veda il progetto ICAR appena citato, <<https://icar.cultura.gov.it/attivita-e-progetti/progetti-icar-1/la-conservazione-del-sito-web-dellistituto-centrale-per-gli-archivi-2008-2016/il-sito-icar-nellinternet-archive>>: «Il primo e più importante servizio che si occupa della raccolta e conservazione di siti web completi è Internet Archive, una organizzazione non-profit che ha lo scopo di realizzare una sorta di archivio di Internet, conservandone i contenuti digitali tramite snapshot cronologici, cioè istantanee delle pagine web, per consentirne l'accesso anche a distanza di tempo, quando tali pagine possono non essere più in linea o essere anche radicalmente cambiate. Gli snapshot sono realizzati attraverso procedure cosiddette di harvesting effettuate da appositi web crawler, che esplorano il web ed effettuano il download delle pagine dei siti web. L'Internet Archive ha sviluppato – e distribuisce con licenze open source – una serie di strumenti e di applicazioni software che hanno favorito la diffusione del modello di raccolta e conservazione di siti web concepito dall'Internet Archive. [...] Tuttavia, analizzando la qualità degli snapshot resi disponibili, si può facilmente constatare una loro pessima corrispondenza con il sito originale sia in termini grafici che di completezza dei contenuti. Inoltre, essendo l'Internet Archive una organizzazione privata, non esiste alcuna garanzia sulla completezza, conservazione e pubblica fruibilità di quanto fino adesso raccolto anche negli anni a venire».

¹⁴ La differenza fondamentale «consiste nel rapporto fra il canale comunicativo (rispettivamente superficie visibile e materia magnetica o simili) e i sensi della persona umana. Mentre il rapporto planare è immediatamente percepibile dal soggetto che scrive o legge, i fenomeni elettromagnetici non sono apprezzabili dei sensi dell'uomo. Il supporto magnetico necessita senza eccezioni di una macchina intermedia fra supporto e lettore»; la comprensione delle differenze tra i due sistemi e dei rispettivi vantaggi e svantaggi «è essenziale per essere consapevoli di ciò che sta avvenendo, ma non deve servire a un'oziosa disputa sulla superiorità dell'uno sull'altro»: T. Orlandi, *Informatica testuale*, cit., p. 15 e 59.

¹⁵ <https://www.agid.gov.it/it/agenzia/strategia-quadro-normativo/codice-amministrazione-digitale>.

«Il tema della dematerializzazione è stato per anni oggetto di lunghi ed a volte aspri dibattiti. Oggi è balzato all'attenzione della Pubblica Amministrazione perché finalmente si vede la concreta possibilità di realizzare quegli ingenti benefici, in termini di risparmio e di efficacia, che il passaggio del documento amministrativo dalla carta al bit ha sempre promesso. Con l'entrata in vigore del Codice dell'amministrazione digitale, viene data attuazione, ricorrendo alle più avanzate tecnologie informatiche, ai meccanismi deputati a realizzare in concreto la *tanto auspicata "scomparsa della carta"* [...] per un percorso concreto di dematerializzazione, da un lato per eliminare i documenti cartacei attualmente esistenti negli archivi, dall'altro per ridurre significativamente la creazione di nuovi documenti cartacei. [...] Per recuperare il passato, cioè per eliminare la *vecchia carta*, occorrerà innanzitutto favorire l'operato delle commissioni di scarto, anche con nuove regole, improntate a pragmatismo e realismo. Nel mondo dei privati, nella maggior parte dei casi è sufficiente solo un po' di iniziativa per ottenere grandi risparmi»¹⁶.

Questa visione dai toni neopositivistici scaturiva per contrasto da una quotidianità di uffici e armadi polverosi, stereotipo della pubblica amministrazione, a cui si contrapponeva la maggiore efficienza dell'imprenditoria privata. L'enfasi sui concetti di "ordine" e "pulizia" appariva strettamente connessa con il "risparmio" previsto, su più piani: nell'ottimizzazione della gestione dello spazio, delle risorse economiche e di quelle umane; portava inoltre con sé una intrinseca connotazione ecologica: a differenza delle vecchie scartoffie, il digitale non consuma carta, inchiostro né toner; non ha bisogno di francobolli né corrieri; se ne deduce che dovrebbe inquinare meno. Tutto questo si traduce nel rinnovato uso del termine "dematerializzazione".

L'espressione "dematerializzazione" si trova già attestata almeno dalla fine dello scorso millennio con il significato di «separazione, eliminazione della parte materiale; scomposizione e vanificazione della materia, della realtà» e deriva da dematerializzare, «privare della materialità; dissolvere la realtà nei suoi elementi costitutivi, privandola di concretezza e consistenza»¹⁷. La parola corrisponde al francese *dématerialiser*¹⁸; sia nella nostra lingua sia in quella francese il prefisso "de", attestato con una gamma di significati che comprendono i concetti di allontanamento, abbassamento, degradazione, fino ad arrivare alla privazione o negazione del ver-

¹⁶ Pierluigi Ridolfi, *Dematerializzazione dei documenti: idee per un percorso*, «I quaderni», 25 (maggio 2006), p. 5-19: la citazione è tratta dall'abstract a p. 5; corsivo mio. Cfr. anche Ministero per l'innovazione e le tecnologie – Gruppo di lavoro per la dematerializzazione della documentazione tramite supporto digitale, *La dematerializzazione della documentazione amministrativa. Libro Bianco del Gruppo di Lavoro interministeriale per la dematerializzazione della documentazione tramite supporto digitale*, Roma: CNIPA, 2006, numero monografico «I quaderni» 24 (aprile 2006), supplemento a «InnovAzione», 9 (2006); e Gianfranco Pontevolpe, *Gli obiettivi del Governo italiano per la dematerializzazione dei documenti*, in: *Conservare il digitale*, a cura di S. Pigliapoco, Macerata: eum, 2010, p. 15-30.

¹⁷ Salvatore Battaglia, *Grande dizionario della lingua italiana*, vol. IV (Dah-Duu), Torino: UTET, 1999, p. 165, disponibile online: <<https://www.gdli.it/JPG/GDLI04/00000172.jpg>>.

¹⁸ *Ibidem*.

bo o sostantivo a cui si antepone, è usato in questo contesto con valore privativo¹⁹. Negli anni successivi il termine è entrato nell'uso comune e figura a buon diritto nel *Lessico del XXI Secolo* Treccani per indicare, «in via generale, il processo e il risultato della *perdita di sostrato materiale e di fisicità*»²⁰.

Se il vocabolo non rappresenta in sé un neologismo, è nuova però la sua applicazione all'ambito amministrativo e archivistico, dove incontra l'ambiente digitale per rappresentare la perfetta sintesi semantica di quanto visto prima. Come opportunamente notato dallo stesso Ridolfi,

«il termine “dematerializzazione” non possiede uno spessore semantico particolare nell'ambito amministrativo, pubblico e privato, ed è totalmente assente nella classica letteratura archivistica, mentre invece compare, anche se da poco tempo, in alcuni testi normativi e all'interno di disposizioni in materia di gestione documentale. Si tratta, tuttavia, di un termine dal forte valore evocativo per identificare la progressiva *perdita di consistenza fisica* da parte degli archivi delle amministrazioni, tradizionalmente costituiti da documenti cartacei, all'atto della loro sostituzione con registrazioni informatiche. È quindi possibile giustificare la nascita di questo termine come conseguenza diretta del progressivo incremento della gestione documentale informatizzata all'interno delle strutture amministrative»²¹.

¹⁹ Cfr. il *Vocabolario on line* Treccani: «de- <dé> [dal lat. *de, de-*]. – 1. Prefisso, soprattutto verbale, che si trova in molte voci di derivazione latina, nelle quali indica ora allontanamento (per es. *deviare, deportare*), ora abbassamento o movimento dall'alto in basso (per es. *degradare, deprimere, declinare*), ora privazione (per es. *dedurre, detrarre*; cfr. anche *demente*), ora ha valore negativo (per es. *decrescere*), ora serve soltanto alla formazione di verbi tratti da sostantivi o aggettivi oppure, con funzione intensiva o con sign. particolari, da altri verbi (per es. *decurtare, designare, determinare*, ecc.). In verbi di formazione recente, spesso formati sull'esempio del francese, corrisponde per lo più a *dis-* o *s-* (per es. *demoralizzare, denaturare*; va notato che in francese il pref. *dé-* corrisponde non all'ital. *de-* ma a *dis-*, essendo derivato dal lat. *dis-*, e indica separazione, privazione o azione contraria); in qualche caso serve a formare doppioni dotti di voci già esistenti (cfr. *demagliare, defogliare, degusciare* di fronte a *smagliare, sfogliare, sgusciare*):»
<<https://www.treccani.it/vocabolario/de/>>.

²⁰ *Lessico del XXI Secolo* Treccani (2012): «dematerializzazione s. f. – Indica, in via generale, il processo e il risultato della perdita di sostrato materiale e di fisicità. Usato prevalentemente in relazione ai documenti di rilevanza giuridica, il termine evoca la trasformazione di un documento cartaceo in un'annotazione, in formato digitale, su di un supporto elettronico che ne consenta – con minori costi e più agevole accessibilità – la formazione, la conservazione, la modificazione e la trasmissione. Il documento dematerializzato che, nel rispetto delle disposizioni normative dettate per ciascun tipo di documento, conserva il valore sostanziale e probatorio di quello cartaceo (o analogico), necessita tuttavia di processi elaborativi di natura informatica, talvolta affidati a soggetti muniti di apposita abilitazione o autorizzazione. Il d. lgs. n. 213 del 24 giugno 1998, recante disposizioni per l'introduzione dell'euro nell'ordinamento interno, ha disposto la d. degli strumenti finanziari negoziati o destinati alla negoziazione sui mercati regolamentati, nonché dei titoli di Stato. Norme sulla d. documentale, sul documento informatico e sulla digitalizzazione delle attività della pubblica amministrazione si trovano oggi raccolte nel Codice dell'amministrazione digitale (v.) così come configurato dal d. lgs. 7 marzo 2005, n. 82 e dalle sue successive modificazioni»:
<[²¹ P. Ridolfi, *Dematerializzazione dei documenti*, cit., p. 6; corsivo mio.](https://www.treccani.it/enciclopedia/dematerializzazione_(Lessico-del-XXI-Secolo)/>; corsivo mio.</p>
</div>
<div data-bbox=)

Nella letteratura di dominio è peraltro attestata anche la variante “smaterializzazione”²², a dimostrazione dell’assimilabilità del significato a un vero e proprio dissolvimento della materia.

Negli anni che sono seguiti, l’espressione è entrata nell’uso comune e gli sforzi si sono concentrati nel superamento dei molti ostacoli concreti che, tra problemi pratici e lungaggini burocratiche, si frappongono quotidianamente alla realizzazione dei piani operativi piuttosto che sulle relative implicazioni terminologiche e definitorie. Nel frattempo però questo ha fatto sì che si consolidasse, inavvertitamente e anche a un livello che non era solo quello degli specialisti, una concezione basata sull’equivoco che il digitale sia qualcosa di completamente astratto e avulso da qualsiasi sostanza fisica, veicolata non solo ma anche dall’uso di espressioni come quelle analizzate sopra. Se una digitalizzazione correttamente impostata può sicuramente contribuire a liberare scrivanie e scaffali, a snellire le procedure e a migliorare la reperibilità dei documenti, il processo rimane, di fatto, quello del trasferimento di un contenuto da un supporto a un altro, per quanto di tipo diverso²³.

Nonostante nei quasi due decenni trascorsi dal varo del CAD abbiamo assistito a un suo progressivo aggiustamento sul piano della teoria e della prassi, il problema dell’impatto ambientale derivante dall’uso delle tecnologie digitali è rimasto in secondo piano fino a tempi molto recenti e stenta ancora a imporsi nella coscienza generale.

4. (Anti)materia ed energia

L’esistenza di un inquinamento derivante dall’uso del digitale è ancora sconosciuta e sottostimata nella coscienza comune, oltre che difficile da calcolare. Tra le attività umane più impattanti sul pianeta, con emissioni di CO₂ equivalente (CO₂e)²⁴ che ammontano a milioni di tonnellate annuali, si collocano le criptovalute, con una stima per il 2019 di 68 milioni di tonnellate²⁵; seguono la tecnologia *in cloud* e i *data centre*, con una stima per il 2020 di 160 milioni di tonnellate; per il

²² Si veda per esempio M. Guercio, *Conservare il digitale: modello nazionale e contesto internazionale*, cit., p. 22.

²³ T. Orlandi, *Informatica testuale*, cit., p. 14-16 e 59.

²⁴ Per una definizione di CO₂ equivalente si veda il glossario online dell’European Environment Agency (EEA): unità utilizzata per misurare le emissioni dei gas serra in base al potenziale di riscaldamento globale (global warming potential – GWP): <<https://www.eea.europa.eu/help/glossary/eea-glossary/carbon-dioxide-equivalent>> (traduzione mia).

²⁵ Si stima che nell’ultimo decennio le criptovalute siano arrivate a rappresentare lo 0,12% dell’impronta ecologica globale. Le criptovalute sono una forma di moneta virtuale che non ha corrispondenza con le monete reali e non si basa su un sistema bancario, ma sull’adozione di una piattaforma condivisa per l’emissione di moneta e la gestione e registrazione delle transazioni. I server costantemente attivi da cui questa dipende sfruttano attualmente circa lo 0,3% dell’energia elettrica mondiale: cfr. Mike Berners-Lee, *How bad are bananas? The carbon footprint of everything*, London: Profile Books, 2020 (revised edition), p. 164-167.

complesso globale delle applicazioni e infrastrutture informatiche il totale raggiunge circa 1.4 bilioni di tonnellate. Un esempio tratto dall'ambito specifico del patrimonio culturale è rappresentato dalla vendita da parte del British Museum di *non-fungible token* (NFT) delle proprie opere, il cui impatto energetico, solo per il periodo settembre 2021-aprile 2022, è assimilabile a quello di un'abitazione nell'arco di circa 57 anni²⁶.

Il dato digitale è basato su flussi di energia che sfruttano supporti magnetici²⁷ e l'evoluzione digitale dipende da un consumo energetico in ascesa: il ricorso sempre più frequente all'intelligenza artificiale produce emissioni circa 10 volte superiori rispetto all'utilizzo di un motore di ricerca di precedente generazione²⁸. L'aumento dell'efficienza rispetto ai precedenti sistemi e la limitazione dell'entropia dell'informazione non sono sufficienti a compensare il maggiore dispendio energetico, in quanto la crescita esponenziale del traffico dati registrata negli ultimi anni rappresenta un perfetto esempio di effetto rimbalzo (*rebound effect*): in seguito all'introduzione di tecnologie più efficienti ed economiche, i benefici del risparmio atteso (nel caso specifico, di natura energetica) sono di fatto vanificati dall'incremento complessivo dei consumi che ne deriva²⁹.

All'apertura del convegno di Europeana 2022 – *making digital culture count*, tenutosi a L'Aia dal 28 al 30 settembre 2022³⁰, Harry Verwayen, General Director

²⁶ Bendor Grosvenor, *The British Museum's NFT project has sent its carbon footprint soaring*, «The Art Newspaper», 5 April 2022: <<https://www.theartnewspaper.com/2022/04/05/nfts-send-carbon-footprints-soaring-british-museum>>. L'articolo è stato citato da Caitlin Southwick nel suo intervento *The Future of Heritage* nell'ambito del convegno di Europeana 2022 – *making digital culture count*, cfr. sotto, nota 39. I *non-fungible token* (NFT) sono oggetti digitali crittografati e inalterabili corredati da un certificato digitale che ne attesta l'originalità e la proprietà univoca: cfr. l'*Enciclopedia* Treccani online: <<https://www.treccani.it/enciclopedia/non-fungible-token/>>.

²⁷ Fabio Ciotti — Gino Roncaglia, *Il mondo digitale: introduzione ai nuovi media*, Roma-Bari: Laterza, 2006, p. 99.

²⁸ John Hennessy, chairman dell'azienda statunitense Alphabet Inc. cui fa capo, tra le altre, anche Google LLC, nel 2023 ha dichiarato che «having an exchange with AI known as a large language model likely cost 10 times more than a standard keyword search»: Jeffrey Dastin — Stephen Nellis, *Focus: For tech giants, AI like Bing and Bard poses billion-dollar search problem*, Reuters, 23 February 2023: <<https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/>>. Si veda anche Alex De Vries, *The growing energy footprint of artificial intelligence*, «Joule», 7 (2023), n. 10, p. 2191-2194, DOI: <<https://doi.org/10.1016/j.joule.2023.09.004>>. Ringrazio Maurizio Lana per i riferimenti.

²⁹ Si parla anche di "paradosso di Jevons", dal nome dell'economista William Stanley Jevons che dedicò il trattato *The coal question; an enquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*, London-Cambridge: MacMillan and Co., 1865, al problema del rapporto apparentemente paradossale tra l'andamento del mercato e l'abbondanza delle risorse: cfr. David Owen, *The Efficiency Dilemma*, «The New Yorker», *Annals of Environmentalism*, 20 & 27 December 2010, disponibile online: <<https://www.newyorker.com/magazine/2010/12/20/the-efficiency-dilemma>>, e M. Berners-Lee, *How bad are bananas?*, cit., p. 166 e 261, nota 17. Cfr. anche Mike Berners-Lee — Duncan Clark, *The burning question: we can't burn half the world's oil, coal and gas. So how do we quit?*, London: Profile Books, 2013.

³⁰ https://pro.europeana.eu/files/Europeana_Professional/Event_documentation/Events/Europeana/2022/Europeana-2022-Digital-Programme.pdf.

della European Foundation, ha ricordato ai presenti che anche in quel momento, sebbene fossero semplicemente seduti in ascolto, stavano contribuendo all'emissione di gas serra con il traffico dati dei propri smartphone, costantemente attivi. Questa considerazione riguarda in effetti la quotidianità della maggior parte di noi e dovrebbe indurci a limitare l'uso della tecnologia non necessaria. In quest'ottica anche la "dematerializzazione", contrariamente agli intenti che hanno guidato la sua genesi e alla propaganda che l'ha sostenuta³¹, rappresenta tutt'altro che una soluzione all'inquinamento: al contrario, anche in ragione del significato che veicola e che si è esaminato prima, dobbiamo chiederci se la "dematerializzazione" non rappresenti oggi piuttosto, in modo più o meno cosciente, un esempio di *greenwashing*³².

5. Riflessioni, spunti e azioni

Un chiaro e documentato prospetto di quali azioni abbiano un impatto ambientale, corredato da strumenti che mettano chiunque in grado di eseguire autonomamente calcoli almeno approssimativi riferiti a qualsiasi attività e di impegnarsi dunque concretamente anche nell'agire, dovrebbe essere presente a tutti per suscitare una nuova consapevolezza e la conseguente rivalutazione di convinzioni maturate in modo acritico³³. Se l'invito alla riduzione volontaria individuale dell'uso del digitale sembra votata a rimanere un'azione velleitaria e dalla scarsa incisività nonostante i lodevoli propositi, il problema deve porsi piuttosto sul piano delle politiche nazionali, teoricamente su scala mondiale.

Se confrontato con la crescita della richiesta di servizi digitali, l'aumento delle

³¹ Cfr. per esempio Nicoletta Pisanu, *PEC, soluzione green: impatto ambientale ridotto e meno carta anche nella PA*, «Agenda Digitale», 13 novembre 2019, online:

<<https://www.agendadigitale.eu/documenti/pec-soluzione-green-impatto-ambientale-ridotto-e-meno-carta-anche-nella-pa/>>, in cui l'utilizzo della posta elettronica certificata (PEC) è incoraggiato per «rendere più ecologiche le procedure documentali, attraverso l'eliminazione della carta» e «ridurre l'impatto ambientale».

³² Cfr. il *Vocabolario on line* Treccani: «Strategia di comunicazione o di marketing perseguita da aziende, istituzioni, enti che presentano come ecosostenibili le proprie attività, cercando di occultarne l'impatto ambientale negativo»:

<https://www.treccani.it/vocabolario/greenwashing_%28Neologismi%29/>.

³³ Un esempio di questi calcoli ragionati costituisce il nucleo dell'opera di Mike Berners-Lee *How bad are bananas? The carbon footprint of everything* (cfr. sopra, nota 25), alla cui prima edizione del 2010 è seguita, dieci anni dopo, una seconda profondamente rivista e aggiornata. In rete si trovano anche degli strumenti per eseguire questi calcoli, per esempio:

<<https://calculator.carbonfootprint.com/calculator.aspx>> e l'EPA Simplified GHG Emissions Calculator, <<https://www.epa.gov/climateleadership/simplified-ghg-emissions-calculator>>. Un altro esempio, relativo alle modalità di viaggio, è rappresentato da Sustainable Travel International, un sito che consente di calcolare l'impatto, in termini di emissioni di gas serra, di uno spostamento differenziato in base al mezzo di trasporto, acquistando eventualmente una compensazione: <<https://sustainabletravel.org/>>. La pratica di compensazione delle emissioni suscita in realtà per lo più diffidenza: cfr. Chris Greenberg, *Carbon offsets are a scam*, Greenpeace, 10 November 2021: <<https://www.greenpeace.org/international/story/50689/carbon-offsets-net-zero-greenwashing-scam/>>.

emissioni di gas serra nel periodo 2015-2021 è da considerarsi, nel complesso, limitato, soprattutto grazie al miglioramento dell'efficienza energetica, al ricorso all'energia rinnovabile e alla decarbonizzazione delle reti elettriche in atto; ciononostante, per realizzare l'obiettivo di azzerare le emissioni entro il 2050 (Net Zero Emissions by 2050 Scenario – NZE), sarebbe necessario arrivare almeno a dimezzarle entro il 2030³⁴.

	2015	2021	Change
Internet users	3 billion	4.9 billion	60%
Internet traffic	0.6 ZB	3.4 ZB	440%
Data centre workloads	180 million	650 million	260%
Data centre energy use (excluding crypto)	200 TWh	220-320 TWh	9.4
Crypto mining energy use	4 TWh	100-140 TWh	+2 300-3 300%
Data transmission network energy use	220 TWh	260-340 TWh	19.4

Tabella 1. *Global trends in digital and energy indicators, 2015-2021 (fonte: Kamiya 2022)*

La campagna di Europeana, la fondazione internazionale sorta per volontà dell'Unione europea e finanziata dalla stessa Unione europea e dagli Stati membri³⁵, per la costruzione di una piattaforma condivisa internazionalmente per i dati del patrimonio culturale digitale si può annoverare tra le iniziative per il riuso delle risorse e dunque per il risparmio energetico legato alla loro alimentazione³⁶. All'interno di Europeana si è inoltre formata recentemente la Climate Action Community, che propone di approfondire la ricerca, portare avanti la divulgazione e impegnarsi pacificamente per la riduzione dell'impatto sull'ambiente delle varie attività umane e in particolar modo di quelle legate al patrimonio culturale, specialmente digitale³⁷. Alla pubblicazione dell'Europeana Climate Action Manifesto nel 2021³⁸ sono seguiti momenti di dialogo e confronto o di informazione e sensibilizzazione; tra questi si segnalano l'incontro *Eco-practices for a green use of digital*, svoltosi online il 9 marzo 2023³⁹, e l'indagine sulla sostenibilità ambientale delle infrastrutture informatiche per la con-

³⁴ George Kamiya, *Data Centres and Data Transmission Networks*, Tracking report – September 2022, International Energy Agency (IEA), <<https://www.iea.org/reports/data-centres-and-data-transmission-networks>>.

³⁵ <https://www.europeana.eu/it/about-us>.

³⁶ Cfr. *Common European data space for cultural heritage*, <<https://pro.europeana.eu/page/common-european-data-space-for-cultural-heritage>>; si veda anche Flavia Bruni, *Europeana Reuse: la Commissione europea e il riuso del patrimonio culturale digitale*, «Digitalia. Rivista del digitale nei beni culturali», sezione monografica *Atti del Convegno di studi "Fare per non sprecare. Nei laboratori del riuso digitale"*, 18, 2 (2023), url open access: <<https://digitalia.cultura.gov.it/article/view/3018/2113>>, p. 147-154: 153.

³⁷ <https://pro.europeana.eu/page/climate-action-community>.

³⁸ <https://pro.europeana.eu/post/the-europeana-climate-action-manifesto>.

³⁹ <https://pro.europeana.eu/event/eco-practices-for-a-green-use-of-digital>.

servazione del patrimonio culturale digitale⁴⁰. I risultati, ancora in corso di elaborazione, dovrebbero contribuire a delineare le pratiche correnti per la gestione e conservazione dei dati nelle istituzioni culturali e consentire dunque di elaborare linee guida strategiche per una trasformazione digitale ecologicamente sostenibile⁴¹. Alla ricerca si accompagnano non solo interventi nell'ambito di congressi e convegni o post divulgativi su *Europeana Pro*⁴², ma anche azioni di "attivismo pacifico", volto soprattutto a spronare i dirigenti a rivedere le politiche pubbliche e aziendali riducendo per quanto possibile gli spostamenti in aereo e il consumo di energia negli uffici per l'illuminazione, il riscaldamento o l'aria condizionata. In occasione del già menzionato convegno di *Europeana 2022 – making digital culture count*, Caitlin Southwick, fondatrice e direttrice di *Ki Culture*, pur premettendo che si tratta di un terreno ancora sostanzialmente inesplorato su cui sono disponibili pochi dati oggettivi («we don't really know»), ha posto alcuni interrogativi sul significato della sostenibilità nelle sue varie implicazioni elencando alcuni tra i principali fattori di impatto ambientale nell'ambito delle professioni del patrimonio culturale digitale⁴³. Per non limitarsi a un approccio teorico, diversi membri dello steering group della *Climate Action Community* hanno

⁴⁰ La ricerca è stata condotta tra gennaio e novembre 2023 da un gruppo di lavoro appositamente costituito comprendente membri dello Steering Group ed esperti esterni: Environmental Sustainability Practice Task Force: <<https://pro.europeana.eu/project/environmental-sustainability-practice-task-force>>. Si veda anche Jose Antonio Gordillo Martorell – Killian Downing, *Is your digital cultural heritage sustainable? Tell us through the Climate Action Community survey*, *Europeana Pro*, 13 July 2023: <<https://pro.europeana.eu/post/is-your-digital-cultural-heritage-sustainable-tell-us-through-the-climate-action-community-survey>>.

⁴¹ *Europeana Climate Action Community Work Plan 2022-2023*, disponibile online: <<https://pro.europeana.eu/post/climate-action-community-work-plan-2022-2023>>, p. 6: «Digital Information Management Survey – design, implementation and analysis. Led by the Environmental sustainability practice Task Force, we will launch a survey on data management and preservation practices in recognition that there is currently no clear overview of how cultural heritage institutions and individuals plan for both a green and digital transformation. We will solicit expert advice if and when needed. This survey is a formative action that supports Europeana's Climate Action Manifesto. It will advocate embedding optimal working practices that minimise the digital cultural heritage sector's impact on the climate and environment. The survey could be followed by identifying and establishing a common vocabulary; clustering of mapping tools and performance indicators, a compilation and comparison of climate action institutional manifestos to develop an informative overview of the sector».

⁴² Cfr. per esempio Shadi Ardalan — Killian Downing, *Europeana Climate Action Community tells us about Earth Day 2023*, *Europeana Pro*, 21 April 2023: <<https://pro.europeana.eu/post/europeana-climate-action-community-tells-us-about-earth-day-2023>>.

⁴³ «*Ki Culture* is a global organization dedicated to creating actionable steps to make culture heritage a leader towards a sustainable future»: <<https://www.kiculture.org/about-us/>>. Riprendendo *Curating tomorrow* di Henry McGhie (<<https://www.curatingtomorrow.co.uk/>>), Caitlin Southwick ha definito sostenibile «doing good without doing harm»; i fattori di impatto da lei identificati sono: «Digitizing collections; Digital footprint: Zoom calls, Storage, Servers, Websites, Social media, Emails; Technology in house; Digital Waste; NFTs; VR/AR; Interactive exhibitions»; ha inoltre più volte rimarcato che il fine non giustifica più i mezzi («Machiavelli is dead») e che non è più possibile trascurare la questione della sostenibilità, neanche per fini ritenuti nobili come quelli culturali. L'intervento, dal titolo *The Future of Heritage*, è integralmente disponibile online: <<https://youtu.be/yX-3rIQikQk?si=8mUHYo0SS7LJK7O3>>.

deciso di raggiungere L'Aia tramite mezzi di trasporto di terra, a discapito di tempo, denaro e comodità, portando con sé coperte, borse dell'acqua calda e giacche pesanti per chiedere di ridurre la temperatura all'interno dell'edificio.

Sebbene con scopi principalmente dimostrativi, sullo sfondo di un panorama generale disperso e frammentario, la presa di posizione da parte di un ente promosso dall'Unione europea appare particolarmente significativa. In ambito nazionale, lo sviluppo ecosostenibile del patrimonio culturale sembra essere ancora poco presente nell'agenda dei dirigenti del Ministero della Cultura (MiC). Le politiche del MiC nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR), programma «di portata e ambizione inedite, che prevede investimenti e riforme per accelerare la transizione ecologica e digitale»⁴⁴, sono riassunte nel Piano nazionale di digitalizzazione del patrimonio culturale (PND) predisposto a partire dal 2022:

«Il Piano nazionale di digitalizzazione del patrimonio culturale (PND) è stato redatto dall'Istituto centrale per la digitalizzazione del patrimonio culturale - *Digital Library* del Ministero della Cultura. Frutto di un processo di condivisione e confronto con diverse istituzioni culturali, il Piano costituisce la visione strategica con la quale il Ministero intende promuovere e organizzare il processo di trasformazione digitale nel quinquennio 2022-2026, rivolgendosi in prima istanza ai musei, agli archivi, alle biblioteche, agli istituti centrali e ai luoghi della cultura statali che possiedono, tutelano, gestiscono e valorizzano beni culturali. Il documento crea il contesto strategico – intellettuale e professionale – di riferimento per la realizzazione degli obiettivi del Piano Nazionale di Ripresa e Resilienza (PNRR), per l'investimento M1C3 1.1 *Strategie e piattaforme digitali per il patrimonio culturale*, e costituisce un utile riferimento metodologico e operativo per tutte le istituzioni e per gli operatori culturali, sia in ambito pubblico che privato, che si riconoscono nei valori in esso enunciati»⁴⁵.

Nonostante «lo scopo [...] di promuovere e orientare il processo di cambiamento degli istituti della cultura verso una trasformazione digitale consapevole, partecipata, condivisa, sostenibile e inclusiva»⁴⁶, l'impegno per uno sviluppo ecosostenibile non è menzionato né tra i valori fondativi né tra gli obiettivi della visione del PND⁴⁷; all'interno di una visione strategica in cui «la pubblica amministrazione ha un ruolo fondamentale nel preservare l'eredità culturale, comunque essa venga declinata, e nel renderla accessibile a tutti nel lungo periodo»⁴⁸, la tematica am-

⁴⁴ Piano Nazionale di Ripresa e Resilienza (PNRR), <<https://italiadomani.gov.it/content/dam/sogeing/documenti/PNRR%20Aggiornato.pdf>>, p. 3.

⁴⁵ <<https://digitallibrary.cultura.gov.it/il-piano/>>; Ministero della Cultura (MiC), Istituto centrale per la digitalizzazione del patrimonio culturale – Digital Library, *Piano Nazionale di Digitalizzazione del patrimonio culturale 2022-2023*, versione n. 1.1 – febbraio 2023: <https://digitallibrary.cultura.gov.it/wp-content/uploads/2023/10/PND_V1_1_2023-1.pdf>, disponibile anche in versione web navigabile selezionando la versione desiderata dalla tendina in alto: <<https://docs.italia.it/italia/icdp/icdp-pnd-docs/it/v1.1-febbraio-2023/index.html>>.

⁴⁶ Ivi, p. 8.

⁴⁷ Ivi, p. 14-22.

⁴⁸ Ivi, p. 14.

bientale si concentra in fugaci accenni alle parole d'ordine «sostenibilità – ecologica, ambientale ed economica»⁴⁹ e transizione verde⁵⁰, concetti di cui manca qualsiasi contestualizzazione e approfondimento, soprattutto per quanto riguarda le implicazioni con il digitale. Neanche le rifiniture apportate al PND in seguito alla consultazione pubblica avvenuta nella primavera del 2022 sono servite a dare spazio alla tematica ambientale⁵¹. L'auspicio non può che essere che i lavori in corso per la digitalizzazione del patrimonio culturale proseguano, in accordo con il contesto internazionale, riservando maggiore attenzione alla tematica ambientale e dimostrino un'accresciuta considerazione delle necessità vitali del pianeta e dei suoi abitanti, umani e non.

Starting from a semantic analysis of the term “dematerialisation”, a key element of the Digital Administration Code (CAD) that came into force on 1 January 2006, and considering other evidence, the article aims to focus on the environmental impact of digital technologies. This issue is also relevant for cultural heritage professionals but still neglected in the recent National Plan for the Digitisation of Cultural Heritage (PND).

⁴⁹ Ivi, p. 40.

⁵⁰ Ivi, p. 5 e 20.

⁵¹ La consultazione sul PND, aperta a tutti, è avvenuta tra il 18 maggio e il 15 giugno 2022 attraverso la compilazione di un questionario e mediante contributi liberi inviati per posta elettronica: cfr. <<https://www.beniculturali.it/comunicato/23192>>, *Esiti della consultazione*, <<https://partecipa.gov.it/processes/piano-nazionale-digitalizzazione-patrimonio-culturale/f/144/>> e *Report finale sulla consultazione*, <https://partecipa.gov.it/uploads/decidim/attachment/file/68/M1C3_1.1.1_ReportFinale_Piano_nazionale_di_digitalizzazione_del_patrimonio_culturale.pdf>, p. 12.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

I requisiti per l'addestramento degli strumenti di Intelligenza Artificiale e il deposito legale delle risorse digitali: una riflessione sul contesto normativo italiano e il ruolo delle biblioteche¹.

«DigItalia» 1-2024

DOI: 10.36181/digitalia-00092

Chiara Storti

Biblioteca nazionale centrale di Firenze (BNCF)

Nel 2022, la Biblioteca nazionale centrale di Firenze ha avviato il progetto sperimentale per l'addestramento in lingua italiana del tool per l'indicizzazione semantica automatica Annif, sviluppato e mantenuto dalla Biblioteca nazionale di Finlandia. La sperimentazione ha fatto emergere una più generale mancanza, nelle biblioteche italiane, dei prerequisiti per l'utilizzo non solo dei più tradizionali strumenti di Machine Learning, ma anche delle nuove Intelligenze Artificiali. Le cause sono da ricondursi, da una parte, alla lacunosa normativa sul deposito legale delle risorse digitali, dall'altra alla frammentazione e moltiplicazione di vocabolari e thesauri, rispondenti ad ontologie differenti e strutturati in formati non sempre adatti all'interoperabilità nel web. Scopo del contributo è quello di avviare una riflessione su questi temi, tentando di delineare il ruolo che potrebbero ricoprire le biblioteche nello sviluppo e utilizzo, etico e consapevole, dei sistemi di Intelligenza Artificiale. Con un approfondimento sul progetto, attualmente in corso, denominato V.I.CO. - Vocabolario delle Identità Controllate.

Premessa

È della fine di dicembre 2023 la notizia che il *New York Times* ha fatto causa alla Microsoft e a OpenAI, la società cui fa capo l'ormai celeberrima ChatGPT,

¹ Il presente contributo riprende in forma approfondita i contenuti dell'intervento dell'autrice dal titolo *Annif per la lingua italiana: i requisiti per l'addestramento degli strumenti di indicizzazione automatica e il progetto V.I.CO*, tenuto in occasione del 62° Congresso AIB, Firenze, 16-17 novembre 2023: <<https://www.aib.it/eventi/congr62/>>. Con l'occasione si rinnovano i ringraziamenti a Giovanni Bergamin, Federico Giubolini e Lorenzo Gobbo, per il fondamentale apporto fornito allo sviluppo dei progetti descritti in questa sede, e ai colleghi dei Settori Ricerche e strumenti di indicizzazione semantica e Servizi informatici della Biblioteca nazionale centrale di Firenze per la generosità nel condividere la documentazione riguardante le prime sperimentazioni sull'indicizzazione semantica automatica condotte in biblioteca, e per gli scambi di idee sul tema sempre arricchenti.

per aver utilizzato gli articoli del quotidiano, coperti da copyright, per addestrare i propri sistemi di Intelligenza Artificiale generativa².

Il tema della titolarità all'utilizzo di dati prodotti da terzi per il training dei sistemi di Intelligenza Artificiale, generativi e non generativi, non è nuovo, ma è entrato anche nel dibattito comune³ da quando, a marzo 2023, OpenAI ha rilasciato l'ultima versione del modello GPT - Generative Pretrained Transformer⁴, consentendo a chiunque di testarne le potenzialità, tramite l'interfaccia di ChatGPT⁵. Questo contributo si pone come una prima e rapida riflessione sulle criticità tecniche e legali, nel contesto italiano, dell'addestramento e dell'uso da parte delle biblioteche degli strumenti di Intelligenza Artificiale, con particolare riferimento a quelli di Natural Language Processing (NLP) e di Machine Learning (ML). Il contributo propone inoltre un approfondimento sui progetti in corso alla Biblioteca nazionale centrale di Firenze (d'ora in poi BNCF) che hanno dato luogo a tali riflessioni.

La prima sperimentazione di indicizzazione semantica automatica dei testi in BNCF

Quasi contestualmente all'avvio, tra il 2010 e il 2011, di Magazzini digitali⁶, il servizio nazionale di conservazione e accesso a lungo termine alle risorse digitali, la BNCF intraprese una sperimentazione per l'indicizzazione semantica automatica delle tesi di dottorato in formato digitale⁷. Il principale strumento utilizzato fu l'indicizzatore MAUI⁸, implementazione basata sull'algoritmo *tf-idf* (*term frequency-inverse document frequency*) che misura l'importanza di un termine all'interno di un testo o di una collezione di testi. I risultati della sperimentazione non furono considerati accettabili rispetto allo standard stabilito, ma iniziarono a

² A titolo esemplificativo, tra le tante testate che hanno riportato la notizia, si veda: *Il New York Times ha fatto causa a OpenAI e Microsoft per aver usato materiale protetto da copyright*, «Il Post», 27 dicembre 2023, <<https://www.ilpost.it/2023/12/27/il-new-york-times-ha-accusato-openai-e-microsoft-di-aver-usato-illecitamente-materiale-protetto-da-copyright/>>.

³ Sempre a titolo esemplificativo, si veda *Il dataset è politico*, «Guerre di Rete - una newsletter di notizie cyber», a cura di C. Frediani, n. 169 (1° ottobre 2023), <<https://guerredirete.substack.com/p/guerre-di-rete-il-dataset-e-politico>>.

⁴ <https://openai.com/gpt-4>.

⁵ Al momento della redazione di questo contributo (gennaio 2024), l'utilizzo di ChatGPT4 è possibile solo a seguito della sottoscrizione di un abbonamento: <<https://chat.openai.com/>>.

⁶ <https://www.bncf.firenze.sbn.it/biblioteca/magazzini-digitali/>.

⁷ Maria Grazia Pepe - Elisabetta Viti, *Prime sperimentazioni di indicizzazione [semi]automatica alla BNCF*, presentazione al 6° incontro ISKO Italia, Firenze, 20 maggio 2013: <<http://www.iskoi.org/doc/firenze13/pepe&viti.pdf>> e Anna Lucarelli - Elisabetta Viti, *Thesaurus del Nuovo soggetto fra linked data, prove di indicizzazione automatica e altri sviluppi*, paper della presentazione tenuta in occasione dell'Italian Research Conference on Digital Libraries (IRCDL), Firenze, 4 febbraio 2016: <https://www.micc.unifi.it/ircdl/wp-content/uploads/2016/01/ircdl2016_paper_12.pdf>.

⁸ MAUI fu sviluppato grazie a un *research grant* di Google: <<https://code.google.com/archive/p/maui-indexer/>>.

far emergere, già in quella sede, la mancanza nel contesto italiano dei prerequisiti per l'addestramento di questo tipo di sistemi. Mancanza di requisiti che, come si vedrà di seguito, persiste ancora oggi.

Strumenti di Intelligenza Artificiale per la valorizzazione e l'accesso alle collezioni digitali e agli archivi del web

Nel 2018, sempre nel perseguimento delle finalità di Magazzini digitali, la BNCF ha inaugurato anche un servizio di *web archiving*⁹, per la conservazione di siti e altre risorse web di interesse culturale nazionale.

A causa della mancanza di uno specifico mandato normativo all'interno della legge sul deposito legale¹⁰, le attività di raccolta periodica, conservazione e accesso sono necessariamente limitate ai soli contenuti per i quali la Biblioteca abbia ricevuto esplicito consenso da parte degli autori o dei responsabili a diverso titolo. La collezione di siti web archiviati dalla BNCF¹¹ è quindi da considerarsi, in termini meramente quantitativi, assai esigua¹². Questa pur piccola collezione contiene, tuttavia, oltre 28 milioni di documenti¹³. Appare evidente come in un ecosistema in cui la quantità di risorse informative che le biblioteche si trovano a gestire sia di tale entità, e sempre in continua crescita, diventi impellente la necessità di individuare strumenti automatici che supportino le tradizionali attività di organizzazione, valorizzazione e accesso alle raccolte.

Da questo punto di vista, i progetti più noti e avviati più precocemente all'interno di biblioteche e archivi riguardano l'utilizzo di sistemi di Natural Language Processing e Machine Learning per il riconoscimento, l'analisi e la metadattazione dei testi¹⁴. Più recenti, invece, i tentativi di mettere a disposizione, in maniera maggiormente efficiente e usabile, le grandi quantità di dati generati dalla creazione di collezioni digitali e degli archivi del web, sia per gli utenti-macchina che

⁹ <https://www.bncf.firenze.sbn.it/biblioteca/web-archiving/>.

¹⁰ Per approfondimenti sul tema, Chiara Storti, *'Resource Not found': Cultural Institutions, Interinstitutional Cooperation and Collaborative Projects for Web Heritage Preservation*, «JLIS.it», 14 (2023), n. 2, p. 39-52, <<https://www.jlis.it/index.php/jlis/article/view/533>>, DOI: 10.36253/jlis.it-533.

¹¹ <https://archive-it.org/home/BNCF>.

¹² A gennaio 2024 i dati raccolti ammontano a circa 3 TB. Per avere un'idea delle dimensioni che possono raggiungere gli archivi del web, è sufficiente citare quello della British Library che, al termine della campagna 2023 di raccolta dello spazio web nazionale, contiene quasi 1.3 Petabyte di dati: <<https://blogs.bl.uk/webarchive/2023/10/uk-web-archive-technical-update-autumn-2023.html>>.

¹³ In questo contesto, "documento" deve intendersi nella definizione data da Archive-it: «a document is any file on the web that has a distinct URL. Images, PDFs, videos, articles, etc., are all considered separate documents».

¹⁴ A questo proposito si veda: Giovanni Colavizza – Tobias Blanke – Charles Jeurgens – Julia Noordegraaf, *Archives and AI: an overview of current debates and future perspectives*, «Journal on Computing and Cultural Heritage», 15 (2021), n. 1, p. 1-15, DOI: <<https://doi.org/10.1145/3479010>>.

per gli utenti-umani. In questa direzione si muovono, da una parte, la ricerca¹⁵ e le sperimentazioni¹⁶ nell'ambito del Computational Access, e dall'altra lo sviluppo di interfacce per il recupero delle informazioni che sfruttino le potenzialità degli strumenti di Intelligenza Artificiale generativa¹⁷.

I risultati del progetto di addestramento del tool Annif per la lingua italiana

Per rispondere all'esigenza di automatizzare, almeno in parte, l'attività di metadattazione delle risorse acquisite nell'ambito del servizio di web archiving, nel 2022 la BNCf ha intrapreso un percorso di valutazione sull'utilizzo per la lingua italiana del toolkit per l'indicizzazione semantica automatica Annif¹⁸, sviluppato e mantenuto dalla Biblioteca nazionale di Finlandia. La riapertura delle attività di ricerca della biblioteca su queste tematiche si deve, infatti, anche alla spinta fornita dall'intervento di Osma Suominen¹⁹ al convegno "Bibliographic Control in the Digital Ecosystem" del febbraio 2021²⁰.

I risultati della sperimentazione, condotta da Lorenzo Gobbo²¹, hanno tracciato una *roadmap* abbastanza precisa dei passi da compiere per l'addestramento di Annif, facendo al contempo emergere, in maniera molto chiara, la mancanza nel contesto nazionale dei prerequisiti di carattere tecnologico, organizzativo e legale per l'adozione da parte delle biblioteche italiane di strumenti di Intelligenza Artificiale, con particolare riferimento a quelli di Machine Learning. Con un azzardato tentativo di generalizzazione, si potrebbe infatti affermare che ciò di cui necessitano sempre questi sistemi sono ampie collezioni di dati e vocabolari strutturati.

¹⁵ Digital Preservation Coalition, *Computational Access: a beginner's guide for digital preservation practitioners*, 2022:

<<https://www.dpconline.org/digipres/implement-digipres/computational-access-guide/>>.

¹⁶ Si veda, ad esempio, l'iniziativa *Computational access to digital collections* della International GLAM Labs Community: <<https://glamlabs.io/computational-access-to-digital-collections/>>.

¹⁷ Si vedano i progetti in corso presso la Library of Congress:

<<https://blogs.loc.gov/thesignal/2023/08/improvements-ahead-for-the-web-archives/>> e presso le Stanford Libraries: <<https://netpreserveblog.wordpress.com/2023/06/28/navigating-through-archived-websites-from-text-matching-to-generative-ai-enhanced-qa/>>.

¹⁸ <<https://annif.org/>>, inizialmente disponibile per il finlandese e l'inglese, è oggi utilizzato anche dalle Biblioteche nazionali di Germania e Svezia per l'indicizzazione dei testi in tedesco e svedese.

¹⁹ Osma Suominen – Juho Inkinen – Mona Lehtinen, *Annif and Finto AI: developing and implementing automated subject indexing*, «JLIS.it», 13 (2022), n. 1, p. 265-282,

<<https://www.jlis.it/index.php/jlis/article/view/437>>, DOI: 10.4403/jlis.it-12740.

²⁰ <https://www.bc2021.unifi.it/>.

²¹ Lorenzo Gobbo allora tirocinante presso la BNCf, ora bibliotecario addetto ai servizi digitali presso USI - Università della Svizzera italiana (lorenzo.gobbo@usi.ch). I risultati sono stati presentati ufficialmente al convegno "Look beyond. Indicizzazione per soggetto delle risorse non librarie", Roma, 6 febbraio 2023:

<<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-studio-catalogazione-ed-indicizzazione/2022/101441-look-beyond-ita/>>. I dati utilizzati per il training di Annif e il risultato della sperimentazione sono disponibili su GitHub: <<https://github.com/Bncf/Annif-Bncf>>.

Annif supporta due diversi formati per il vocabolario di riferimento: una versione SKOS²², ontologia basata su RDF per la rappresentazione delle relazioni semantiche tra entità all'interno di thesauri, vocabolari e sistemi di classificazione, oppure un più semplice formato testuale TSV in cui i dati sono rappresentati come coppie di URI e etichette alternative e letterali.

Di seguito un esempio dell'entità "pubblicazioni scientifiche"²³ con le sue relazioni, come descritte all'interno del Thesaurus del *Nuovo soggettario*, in formato SKOS:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:skos-xl="http://www.w3.org/2008/05/skos-xl#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:iso-thes="http://purl.org/iso25964/skos-thes#"
xmlns:void="http://rdfs.org/ns/void#"
xmlns:nsogi="http://prefix.cc/nsogi">

<rdf:Description rdf:about="http://purl.org/bnfc/tid/10042">
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
<skos:prefLabel xml:lang="it">Pubblicazioni scientifiche</skos:prefLabel>
<dc:date>2005-06-17</dc:date>
<skos:inScheme rdf:resource="http://purl.org/bnfc/tid/ThesCF15"/>
<skos:inScheme rdf:resource="http://purl.org/bnfc/tid/Thes"/>
<skos:broader rdf:resource="http://purl.org/bnfc/tid/584"/>
<skos:narrower rdf:resource="http://purl.org/bnfc/tid/33325"/>
<skos:narrower rdf:resource="http://purl.org/bnfc/tid/4473"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/57934"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/9070"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/17374"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/50001"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/2762"/>
<skos:related rdf:resource="http://purl.org/bnfc/tid/1977"/>
<skos:closeMatch
rdf:resource="http://www.wikidata.org/wiki/Q591041"/><skos:editorialNote>FONTE:
Soggettario; DeM: scientifico; VT; Wikipedia(IT)</skos:editorialNote>
</rdf:Description>
</rdf:RDF>
```

E un esempio di lista di entità, prive di relazioni esplicite, in formato TSV:

²² <https://www.w3.org/2004/02/skos/intro>.

²³ <https://thes.bnfc.firenze.sbn.it/termine.php?id=10042>.

<<http://purl.org/bnctf/tid/10041>> Pubblicazioni ufficiali
 <<http://purl.org/bnctf/tid/10042>> Pubblicazioni scientifiche
 <<http://purl.org/bnctf/tid/10043>> Pubblicazioni militari

Non tutti gli algoritmi per l'indicizzazione, però, supportano o hanno le medesime performance con entrambi i formati²⁴, rivelando invece l'opportunità di avere a disposizione per l'addestramento un vocabolario SKOS.

Il vocabolario, inoltre, deve essere abbastanza ampio da referenziare tutti i termini usati per l'indicizzazione del *corpus* di addestramento. In Italia gli authority file tradizionalmente implementati e usati nel dominio GLAM/MAB sono gestiti in archivi separati con strutture e formati di dati spesso molto differenti.

Condizione necessaria, quindi, per l'addestramento di Annif ai fini di un suo utilizzo sistematico e non solo sperimentale, è l'individuazione di una soluzione per la gestione integrata dei dati di autorità in formati standard e adatti all'interoperabilità nel web, sul modello del tedesco GND - Gemeinsame Normdatei²⁵.

Allo stesso tempo, il *corpus* di addestramento deve essere costituito da porzioni di testo sufficientemente lunghe e "significative", capaci cioè di rappresentare adeguatamente il contenuto precipuo della risorsa informativa, abbinata a liste di termini riconducibili alle entità presenti nel vocabolario, che ne identifichino il contenuto semantico. Perché l'addestramento sia completo, i diversi ambiti disciplinari dovrebbero essere rappresentati in maniera omogenea. Purtroppo, anche questa seconda condizione, a causa dei limiti normativi e organizzativi di cui soffrono le biblioteche e le istituzioni culturali nel nostro Paese, non è di facile assolvimento.

Il deposito legale delle risorse digitali per la costituzione di corpora di dati affidabili e verificabili²⁶

Dal punto di vista delle biblioteche, tra le principali novità introdotte nella normativa italiana sul diritto d'autore – Legge 22 aprile 1941, n. 633 –, a seguito del recepimento nel 2021 della Direttiva UE 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale, si annoverano quelle presenti nel Capo V del Titolo 1 "Eccezioni e limitazioni", che possono essere sintetizzate nell'enunciato dei commi 1 e 2 dell'art. 70 ter.:

«1. Sono consentite le riproduzioni compiute da organismi di ricerca e da istituti di tutela del patrimonio culturale, per scopi di ricerca scientifica, ai fini dell'estrazione di testo e di dati da opere o da altri materiali disponibili in reti o banche di dati cui essi hanno lecitamente accesso, nonché la comunicazione al pubblico degli esiti della ricerca ove espressi in nuove opere originali.

²⁴ Soltanto tre dei dieci algoritmi testati possono essere usati senza un vocabolario in formato SKOS.

²⁵ https://gnd.network/Webs/gnd/DE/Home/home_node.html.

²⁶ Su questo tema verteva anche l'intervento al 62° Congresso AIB di Elda Merenda, dal titolo *Deposito legale e intelligenza artificiale*: <<https://www.aib.it/eventi/congr62/#capitolo-2>>.

2. Ai fini della presente legge per estrazione di testo e di dati si intende qualsiasi tecnica automatizzata volta ad analizzare grandi quantità di testi, suoni, immagini, dati o metadati in formato digitale con lo scopo di generare informazioni, inclusi modelli, tendenze e correlazioni.».

Precedentemente, il Regolamento UE 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati (GDPR) aveva sancito la possibilità di conservare anche i dati personali «a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici», come eccezione alla norma generale sul diritto all'oblio²⁷, stabilendo contemporaneamente le modalità del trattamento. Dal combinato disposto di queste due norme si evince chiaramente che le biblioteche, come gli altri istituti di tutela del patrimonio e gli organismi di ricerca, potrebbero utilizzare le collezioni di testi digitali e, più in generale, i *dataset* di cui sono legittimamente entrati in possesso, per lo sviluppo e il training di sistemi di Machine Learning.

Il nodo, potremmo affermare, si situa proprio nelle *legittime* modalità di acquisizione dei dati. Infatti, tradizionalmente, il canale preferenziale per l'accrescimento del patrimonio delle biblioteche è l'istituto del deposito legale²⁸ che prevede, in maniera simile in quasi tutto il mondo, l'obbligo di consegna presso le istituzioni individuate di tutti i documenti di interesse culturale pubblicati sul territorio nazionale. Anche in Italia, come nella maggioranza degli altri paesi europei, tale obbligo era stato esteso formalmente alle "risorse diffuse tramite rete informatica", vale a dire alle risorse nativamente digitali e alle risorse web, già con il d.P.R. 3/5/2006, n. 252. Tuttavia, esso non è mai stato regolamentato dal punto di vista tecnico, così come previsto nell'articolo 37²⁹ dello stesso d.P.R. 3/5/2006, n. 252, ed è quindi rimasto inattuato³⁰. A causa di questa grave lacuna normativa, le risorse native digitali e web conservate nelle biblioteche, come si è già avuto modo di sottolineare in relazione al *web archiving*, sono quantitativamente ridotte, venen-

²⁷ Regolamento UE 2016/679 art. 17.

²⁸ La vigente normativa italiana sul deposito legale è riconducibile sostanzialmente alla l. 106/2004 e al d.P.R. 3/5/2006, n. 252. Maggiori approfondimenti sono disponibili sul sito della Direzione generale biblioteche e diritto d'autore: <<https://biblioteche.cultura.gov.it/it/Attivita/deposito-legale/>>.

²⁹ L'art. 37 c. 1 del d.P.R. 3/5/2006, n. 252 così recita: «Le modalità di deposito dei documenti diffusi tramite rete informatica sono definite con successivo regolamento adottato ai sensi dell'articolo 5, comma 1, della l. 106/2004, su proposta del Ministro per i beni e le attività culturali, di concerto con il Ministro delegato per l'innovazione e le tecnologie, sentite le associazioni di categoria interessate, nonché la Commissione per il deposito legale, di cui all'articolo 42 e il Comitato consultivo permanente per il diritto d'autore».

³⁰ Tra i più recenti contributi sul tema si veda Paola Puglisi, *Deposito legale quattordici anni dopo: come, quando, 'quanto', e perché*, «AIB Studi», 60 (2020), n. 3, <<https://doi.org/10.2426/aibstudi-12477>> e Giuliano Genetasio – Elda Merenda – Chiara Storti, *In the mangrove society: a collaborative legal deposit management hypothesis for the preservation of and permanent access to the National Cultural Heritage*, «JLIS.it», 13 (2022), n. 1, <<https://doi.org/10.4403/jlis.it-12732>>.

do acquisite principalmente grazie a depositi di tipo volontario o a seguito di esplicita richiesta di autorizzazione ai titolari. L'assenza di un mandato normativo specifico implica anche la generale grave carenza negli istituti del patrimonio di adeguate risorse, sia professionali che economiche, che sarebbero necessarie per un'efficiente gestione del patrimonio digitale. Si può quindi affermare che le biblioteche oggi non dispongono, e difficilmente potrebbero disporre, *legittimamente* di collezioni di testi digitali, in quantità utili all'addestramento di strumenti come Annif o di altri modelli di Natural Language Processing³¹.

Ed è doveroso sottolineare che non si perde così solo l'opportunità che i sistemi di Intelligenza Artificiale servano davvero a migliorare i servizi delle biblioteche, ma anche e soprattutto che tali sistemi siano addestrati in maniera etica, garantendo totale trasparenza sulla fonte di acquisizione e le modalità di trattamento dei dati, oltre che la possibilità di condivisione dei modelli così ottenuti.

Rispetto alla disponibilità di collezioni di testi adatte al training degli strumenti di NLP e ML, vi è un ulteriore limite che potremmo definire "intrinseco" all'organizzazione degli attuali flussi catalografici nelle biblioteche italiane, in parte ancora conseguenza dall'assenza di obbligo di deposito legale delle risorse native digitali. In Italia, infatti, le risorse che rientrano nei tradizionali meccanismi di controllo bibliografico³² e che sono soggette a procedure di indicizzazione semantica "controllata" – utili quindi per il training –, non sono disponibili in versione digitale, almeno non in modalità open³³. Viceversa, le risorse maggiormente o esclusivamente disponibili in digitale, spesso in open access e dotate di ricchi metadati, sono relative a campi del sapere che in gran parte non rientrano nel controllo bibliografico strettamente inteso. Tra queste sicuramente la quasi totalità dei prodotti della ricerca, in ambito scientifico, che restano quasi esclusivo appannaggio dei gestori dei repository accademici. Per questi motivi, per la creazione di *corpora* di testi in grado di rappresentare molteplici campi del sapere, sarebbe necessario stabilire delle forme di cooperazione con le biblioteche e i sistemi bibliotecari di ateneo, oltre che con editori e distributori specializzati e no³⁴.

³¹ Per farsi un'idea delle potenzialità dei *dataset* conservati dalle biblioteche per il NLP: *The Spanish Web Archive as a training field for Natural Language Processing models*, «IIPC Blog», 29 settembre 2021: <<https://netpreserveblog.wordpress.com/2021/09/29/the-spanish-web-archive-as-a-training-field-for-natural-language-processing-models/>>.

³² Si fa qui riferimento, in primo luogo, alla redazione della Bibliografia nazionale italiana: <<https://www.bncf.firenze.sbn.it/biblioteca/bibliografia-nazionale-italiana/>>.

³³ Nel contesto appena delineato relativo alla mancanza di obbligo di deposito legale delle risorse digitali, con la conseguente possibilità di effettuare legittimamente attività di *text and data mining*, le licenze aperte sono requisito indispensabile. È però doveroso precisare che la sperimentazione ha dimostrato che sono sufficienti anche porzioni testuali più ridotte, come gli abstract, e che questi si potrebbero ottenere ad esempio con un accordo con i gestori del database dei libri in commercio.

³⁴ La creazione, gestione e conservazione a lungo termine di collezioni di dati, anche per finalità come quelle qui delineate, nell'ambito della cultura e della ricerca scientifica, sarebbe facilitata dalla creazione di una rete di soggetti sia pubblici che privati come quella ipotizzata in Associazione italiana

V.I.CO. - Vocabolario delle Identità Controllate

Il progetto V.I.CO. - Vocabolario delle Identità Controllate³⁵, avviato sempre in BNCf nel 2023, ha lo scopo di mettere a disposizione, in un ambiente nativamente linked open data, un vocabolario controllato partecipato dei termini in lingua italiana relativi alle entità (concetti, persone e organizzazioni, opere, luoghi, eventi ecc.) utili alla descrizione del contenuto dei testi, ovvero alla loro indicizzazione descrittiva e semantica. Come si è già avuto modo di rilevare, la disponibilità di un vocabolario in formato aperto è, infatti, uno dei requisiti per l'avvio dell'indicizzazione automatica dei testi in lingua italiana tramite il toolkit Annif.

In Italia i vocabolari controllati, incluse le liste di autorità, tradizionalmente implementati e usati per la descrizione e l'indicizzazione delle risorse nel dominio della cultura, sono gestiti in archivi separati, rispondenti a differenti ontologie. Un elenco non esaustivo comprende:

- il Thesaurus³⁶ del *Nuovo soggettario*;
- l'edizione italiana della Classificazione Decimale Dewey³⁷;
- gli authority file gestiti dall'ICCU relativi a Nomi, Opere e Luoghi³⁸;
- le liste di autorità prodotte in ambito archivistico³⁹;
- i vocabolari prodotti e mantenuti dall'ICCD⁴⁰.

A questi si potrebbero aggiungere i vocabolari prodotti al di fuori dell'ambito GLAM/MAB, in particolare nel dominio della ricerca, e ugualmente utili all'indicizzazione semantica delle risorse. A titolo esemplificativo:

- gli elementi Wikidata che presentano una *label* in italiano⁴¹;
- il database dei nomi geografici GeoNames⁴²;
- le liste di nomi di autori presenti nei repository della ricerca⁴³;
- l'elenco dei settori scientifico-disciplinari (SSD) del MUR.

biblioteche, Gruppo di lavoro sulle biblioteche digitali, *Piano d'azione per l'infrastruttura nazionale della conoscenza*. Roma: Associazione Italiana Biblioteche, 2023. ISBN 978-88-7812-374-8, DOI: <<https://doi.org/10.53263/9788878123748>>.

³⁵ https://vicobncf.wikibase.cloud/wiki/Main_Page.

³⁶ <https://thes.bncf.firenze.sbn.it/>.

³⁷ Attualmente l'edizione italiana della CDD ed. 23. o *Web Dewey* non è disponibile con licenza e formati aperti. A questo proposito si confrontino il contributo di Pietro Cavaleri, *WebDewey search*, «AIB Studi», 57 (2017), n. 2, <<https://doi.org/10.2426/aibstudi-11644>> e il progetto tedesco Web Dewey Search: <<https://deweysearchde.pansoft.de/>>.

³⁸ <https://www.iccu.sbn.it/it/normative-standard/authority-control/>.

³⁹ <https://icar.cultura.gov.it/standard/standard-san>.

⁴⁰ <http://www.iccd.beniculturali.it/it/strumenti-terminologici>.

⁴¹ La query SPARQL effettuata sul database di Wikidata, tramite il servizio QLever, restituisce oltre 3.770.000 risultati (dati a gennaio 2024): <<https://qllever.cs.uni-freiburg.de/wikidata/ewLSC3>>.

⁴² <https://www.geonames.org/>.

⁴³ A questo proposito, di particolare interesse il progetto IRIS del Gruppo Wikidata per Musei, Archivi e

V.I.CO. è stato quindi pensato come soluzione per la gestione condivisa di vocabolari controllati, alimentato da dati provenienti da diversi sistemi di dominio, sia centrali che locali. È ospitato su una istanza Wikibase Cloud⁴⁴. La disponibilità di Wikibase come SaaS⁴⁵ consente di concentrare le limitate risorse a disposizione sulla analisi e gestione dei dati più che sulle necessità sistemistiche. Inoltre, da qualche anno, Wikibase è utilizzato a livello globale da numerose istituzioni MAB/GLAM per la gestione dei dati di dominio⁴⁶. In particolare, è in rapida crescita il numero di biblioteche nazionali europee che lo adottano per l'allestimento di vocabolari controllati, in maniera esclusiva o, più frequentemente, come "repository LOD" dei dati. Tra queste, le biblioteche nazionali francese⁴⁷, tedesca⁴⁸, ceca⁴⁹, greca e portoghese.

Per il popolamento di V.I.CO. sono stati, innanzitutto, individuati i vocabolari di interesse prioritario rispetto all'utilizzo dei termini in essi contenuti per l'indicizzazione dei testi delle collezioni a disposizione per il training, e il loro ordine di importazione nell'istanza Wikibase⁵⁰.

In questa fase si è scelto di non riconciliare esplicitamente eventuali entità descritte in più di un vocabolario, come ad esempio le entità persona presenti sia in Wikidata che nell'Authority file Nomi dell'ICCU, ma per ciascuna entità sono stati importati, laddove presenti, gli identificativi dei vocabolari esterni, per garantire comunque la possibilità di riconciliazione lato macchina⁵¹. Per ciascuna tipologia di entità (nomi, luoghi ecc.), e per ciascun vocabolario, è stato poi definito un *subset* di dati di interesse ai fini dell'indicizzazione con Annif. Limitando in questo modo

Biblioteche:

<https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/IRIS>, presentato anche in occasione della Giornata di studi dal titolo "La valutazione della ricerca alla prova dei fatti: modelli, politiche e progetti in corso", organizzata dall'Osservatorio della ricerca dell'Università di Firenze e dal Sistema bibliotecario di Ateneo, in collaborazione con Wikimedia Italia, Firenze, 17 gennaio 2024.

⁴⁴ <<https://www.wikibase.cloud/>>. Wikibase è un software open source per la creazione di basi di dati collaborative rispondenti allo standard RDF, nativamente multi-ontologie. Il software è mantenuto dall'associazione no profit Wikimedia Deutschland e da un'ampia comunità globale:

<<https://wikiba.se/>>. La sua implementazione più nota è Wikidata, database libero e collaborativo contenente oltre 100 milioni di dati: <<https://www.wikidata.org/>>.

⁴⁵ SaaS – *Software as a Service* indica la modalità di distribuzione di un software o programma: il software è ospitato su una infrastruttura cloud ed è reso disponibile agli utenti attraverso un semplice client, come un web browser. Si veda la definizione fornita dal National Institute of Standards and Technology (NIST): <https://csrc.nist.gov/glossary/term/software_as_a_service>.

⁴⁶ <https://learning.wikibase.com/usecases/>.

⁴⁷ <https://www.transition-bibliographique.fr/fne/fichier-national-entites/>.

⁴⁸ <<https://wiki.dnb.de/pages/viewpage.action?pageId=147754828>> e <https://docs.google.com/presentation/d/1_xdkofNladhGJsNnINgovqsz2UHJSVwpPkc3dpXeUs>.

⁴⁹ https://autority.wikimedia.cz/auto/Hlavn%C3%AD_strana.

⁵⁰ Per la scelta dei vocabolari si veda l'homepage di V.I.CO.:

<https://vicobncf.wikibase.cloud/wiki/Main_Page>.

⁵¹ A titolo esemplificativo è possibile confrontare l'entità riferita all'autore Carlo Collodi descritta nativamente nell'authority file dell'ICCU: <<https://vicobncf.wikibase.cloud/wiki/Item:Q17>>, e quella rappresentante la medesima entità in Wikidata: <<https://vicobncf.wikibase.cloud/wiki/Item:Q19>>.

la quantità di dati da importare in V.I.CO. e gli eventuali conseguenti problemi sia di carattere sistemistico che applicativo⁵²; ferma restando l'attribuzione ad ogni entità di una proprietà relativa alla fonte originale dei dati, così da facilitare un possibile necessario ampliamento del tracciato dati e, più in generale, il loro aggiornamento. Successivamente sono state mappate le ontologie dei dataset originali ed è stato implementato il meta-modello dati di V.I.CO⁵³. È bene specificare che il modello dati di V.I.CO. non ha ambizioni di tipo "ontologico", ma è puramente funzionale alla gestione in unico ambiente di vocabolari strutturati in base a logiche di dominio differenti. Infine, sono state sperimentate le modalità di import di ciascun dataset in V.I.CO., dipendenti dai formati e dalle possibilità di export dei vocabolari di origine.

A questo proposito una delle principali criticità da risolvere per la messa in produzione del servizio è la definizione dei criteri e delle modalità di aggiornamento dei dati nell'istanza Wikibase, sia in relazione all'aggiornamento dei dataset-fonte, sia rispetto alla possibilità di introdurre direttamente in V.I.CO. entità utili all'indicizzazione semantica e non già descritte altrove.

Appare evidente, in conclusione, come un progetto di tale portata richieda l'apporto di altre istituzioni e organizzazioni nazionali, non solo in qualità di fornitrici di dati ma, soprattutto, come portatrici di competenze sui temi in oggetto. La finalizzazione di accordi di collaborazione con i principali attori nazionali in questo ambito⁵⁴, tutti purtroppo affetti da gravi carenze di personale, si configura come una ulteriore sfida, questa volta di carattere amministrativo e organizzativo, per l'utilizzo dei sistemi di Intelligenza Artificiale nelle biblioteche.

Ampliando la prospettiva: Annif, Semantic Search e IA generativa

Il 2023 è stato, senza dubbio, l'anno dell'Intelligenza Artificiale generativa. Sistemi come ChatGPT4 rappresentano, per ora, uno dei punti di arrivo più raffinati e allo stesso tempo più "pop" della generazione degli strumenti che utilizzano modelli linguistici (LLM) creati grazie all'impiego delle reti neurali profonde con un'architettura basata sui *transformer*⁵⁵. Tali reti neurali profonde prendono in considera-

⁵² In generale non è ignorabile la necessità di disporre di risorse tecnologiche e computazionali per la gestione di grandi dataset.

⁵³ Classi e proprietà generali ontologia VICO: <<https://tinyurl.com/36z3ukht>>.

⁵⁴ Tra questi, in prima istanza, gli istituti del MiC deputati alla definizione degli standard nazionali di descrizione e metadattazione e alla gestione degli authority file ovvero ICCU, ICAR e ICCD, ma anche l'ICDP-DL che, nell'ambito delle attività previste dall'investimento "M1C3 1.1 Strategie e piattaforme digitali per il patrimonio culturale" del Piano nazionale di ripresa e resilienza (PNRR), sta lavorando alla costituzione di una infrastruttura nazionale dei dati della cultura. Inoltre, colloqui informali sono stati avviati con il gruppo GWMAB che ha ormai un'esperienza pluriennale nella gestione di dati in ambiente Wikibase/Wikidata: <https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche>.

⁵⁵ Un'efficace rappresentazione del funzionamento dei *transformer* è quella fornita nell'articolo *Generative AI exists because of the transformer*, pubblicato il 12 settembre 2023 sul «Financial

zione un grandissimo numero di parametri ma, soprattutto, sfruttano una modalità di analisi dei *token* linguistici che potremmo definire altamente “semantica”, poiché li esamina non soltanto in base alla loro sequenza ma ne soppesa il contesto d’uso. Queste tecniche di apprendimento sono generalmente non supervisionate, cioè non necessitano di intervento umano per l’etichettatura dei dati, almeno non in una prima fase⁵⁶. Il risultato sono interfacce di *Semantic Search*⁵⁷ e di IA generativa le cui capacità di recupero dell’informazione, anche in presenza di enormi quantità di dati, sono evidentemente sorprendenti, e sempre meno affette da “allucinazioni”⁵⁸. In questo contesto, ci si potrebbe chiedere quale sia il senso di continuare a investire sullo sviluppo di strumenti di indicizzazione semantica automatica⁵⁹, come Annif, basati su forme tradizionali di Machine Learning. Utilizzando dati pre-addestrati secondo un modello tradizionale di organizzazione della conoscenza, Annif sembra rispondere alla sola esigenza di velocizzare e rendere maggiormente efficienti processi consolidati che vedono il bibliotecario (o un qualsiasi esperto in un dato ambito del sapere) tentare di prevedere non solo *cosa* gli utenti cercheranno, ma anche *come* lo faranno, utilizzando una sintassi e una semantica che sono appannaggio quasi esclusivo del mondo MAB/GLAM.

La *Semantic Search* e l’AI generativa ribaltano, invece, la prospettiva cercando di restituire risultati quanto più *probabilmente* attinenti alla ricerca dell’utente, indipendentemente dalla sua formulazione. Questi sistemi, infatti, forniscono risposte proprio sulla base di modelli probabilistici complessi creati, come appena accennato, grazie all’utilizzo di reti neurali profonde che analizzano grandi quantità di dati. Ci troviamo cioè davanti a quell’apparente dualismo nei sistemi di organizzazione della conoscenza che Gino Roncaglia ha di recente egregiamente sintetizzato nella metafora dell’Architetto e dell’Oracolo⁶⁰. Come affermato anche dallo stesso Roncaglia, non siamo ora in grado di dire se e quale approccio prevarrà sull’altro, e con che tempi. Restano però valide, qualsiasi sia lo scenario futuro, tutte le considerazioni circa la necessità di avere finalmente una normativa nazionale che preveda l’obbligo di deposito legale delle risorse native digitali, che consenta alle biblioteche di essere parte attiva nella costruzione di tali strumenti. Allo stesso tem-

Times», a firma del Visual Storytelling Team e di Madhumita Murgia:

<<https://ig.ft.com/generative-ai/>>.

⁵⁶ I modelli ottenuti partendo da un processo di apprendimento autonomo possono essere perfezionati tramite processi di apprendimento supervisionato.

⁵⁷ <https://www.sbert.net/examples/applications/semantic-search/README.html>.

⁵⁸ Andrea Fontana, *Anche l’intelligenza artificiale può avere le allucinazioni*, «Wired.it», 18 giugno 2023: <<https://www.wired.it/article/intelligenza-artificiale-allucinazioni-cause-conseguenze/>>.

⁵⁹ Interessante la serie di progetti e soluzioni presentate in «Cataloging & Classification Quarterly», 59 (2021), n. 8, fascicolo interamente dedicato ai sistemi di indicizzazione semantica automatica: <<https://www.tandfonline.com/toc/wccq20/59/8>>.

⁶⁰ Gino Roncaglia, *L’architetto e l’oracolo: forme digitali del sapere da Wikipedia a ChatGPT*, Bari; Roma: Laterza, 2023.

po, all'interno delle istituzioni culturali si dovrebbe avviare una riflessione più profonda sul concetto di *qualità* dei dati prodotti, che non può più essere misurata solo in base all'aderenza a standard stabiliti all'interno delle stesse comunità di dominio, ma che dovrebbe piuttosto considerare la possibilità di un loro riuso anche per l'addestramento dei sistemi di Intelligenza Artificiale. Se non si invertono le attuali tendenze, le biblioteche in Italia non saranno in grado di utilizzare le reali potenzialità delle Intelligenze Artificiali, e soprattutto non potranno esercitare il ruolo di produttori e gestori di conoscenza secondo i principi, che sono loro propri, di legalità, eticità e trasparenza.

In 2022, the National Central Library of Florence launched an experimental project to train the Annif tool for automatic semantic indexing in the Italian language, developed and maintained by the National Library of Finland. The project revealed a more general lack of the prerequisites in Italian libraries for using not only the more traditional Machine Learning tools but also new Artificial Intelligence systems. The causes can be traced, on the one hand, to deficient legislation on the legal deposit of digital resources and, on the other hand, to the fragmentation and multiplication of vocabularies and thesauri, which respond to different ontologies and are structured in formats not always suitable for web interoperability. The purpose of this contribution is to initiate a reflection on these issues, attempting to outline the role that libraries could play in the ethical and conscious development and use of Artificial Intelligence systems, with an in-depth look at the ongoing project called V.I.CO. - Vocabulary of Controlled Identities.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Il riversamento di formato elettronico tra standard internazionali e Linee guida sulla formazione, gestione e conservazione di documenti informatici dell'AgID

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00093

Stefano Allegrezza

Università degli Studi di Bologna

Il problema dell'obsolescenza dei formati elettronici implica sfide significative per la conservazione a lungo termine dei documenti digitali. I continui sviluppi tecnologici rendono obsoleti i formati esistenti, richiedendo la conversione tempestiva in nuovi formati per garantirne la fruibilità futura. Questo problema riguarda sia i formati open source che quelli commerciali, con fornitori che talvolta incoraggiano il passaggio a nuovi prodotti attraverso l'obsolescenza programmata. Per risolvere tale problema è stata proposta la strategia del riversamento di formato elettronico, che comporta la conversione dei documenti digitali dal vecchio formato ad un formato più attuale. Tuttavia, nonostante l'importanza di questa strategia sia ampiamente riconosciuta, a livello pratico la sua attuazione è ancora limitata a causa della mancanza di conoscenze e competenze, di linee guida e raccomandazioni operative. Questo contributo mira a colmare questa lacuna, delineando le motivazioni alla base del riversamento di formato, il momento in cui attuarla e le metodologie alla base della conversione – anche con riferimento alle linee guida internazionali e nazionali – e proponendo una metodologia per la conversione massiva basata sulla certificazione di processo.

1. Introduzione

Il problema dell'obsolescenza dei formati elettronici rappresenta una grande sfida per la conservazione a lungo termine, non solo per gli archivi e le biblioteche digitali ma per qualsiasi repository di oggetti digitali. Infatti, questi ultimi, indipendentemente dalla categoria di appartenenza (documenti testuali, immagini, documenti sonori e audiovisivi, elaborati tecnici ecc.) e dal formato elettronico in cui sono codificati, sono destinati prima o poi a diventare obsoleti (e in alcuni casi lo sono già). Ciò avviene perché man mano che vengono sviluppate nuove funzionalità, nuovi formati (o nuove versioni di formati già esistenti) vengono introdotti sul mercato. In questo modo, i vecchi formati (o le vecchie versioni dei formati) di-

ventano obsoleti e nelle nuove generazioni di software viene gradualmente eliminato il supporto per tali formati. La conseguenza di tutto ciò è che nel tempo diventa sempre più difficile visualizzare o riprodurre i documenti codificati secondo tali formati obsoleti¹. Si noti che questo problema riguarda tanto i formati commerciali che quelli open source: da una parte le aziende sviluppatrici di software utilizzano l'obsolescenza programmata per invogliare i clienti a passare ai nuovi prodotti e conseguentemente ai nuovi formati; dall'altra le comunità open source ritirano il supporto per i formati più vecchi quando questi si rivelano sempre meno utilizzati o non sono più necessari alla comunità. L'obsolescenza può anche essere accidentale: sia le aziende che le comunità open source possono fallire.

La soluzione al problema dell'obsolescenza dei formati elettronici è stata individuata nella strategia del riversamento di formato, che consiste nella conversione di un oggetto digitale da un formato fruibile in un determinato ambiente hardware e software in un formato fruibile in un altro ambiente mantenendone le proprietà significative² (Fig. 1). Tipicamente viene messa in atto quando il formato elettronico di un oggetto digitale comincia ad essere obsoleto, ma ci possono essere anche motivazioni di altra natura. Ad ogni modo, il riversamento di formato deve essere eseguito tempestivamente quando i cambiamenti nell'ambiente hardware e software cominciano a minacciare la possibilità di fruizione di un determinato formato elettronico. Non ammette dilazioni e deve essere completato prima che il formato corrente divenga del tutto obsoleto. È un processo critico nei confronti del tempo: se venisse persa una generazione di nuovi formati, gli oggetti digitali potrebbero essere difficili da recuperare; se venissero perse più generazioni, gli oggetti digitali potrebbero addirittura essere irrecuperabili. I cicli di conversione necessitano di essere relati-

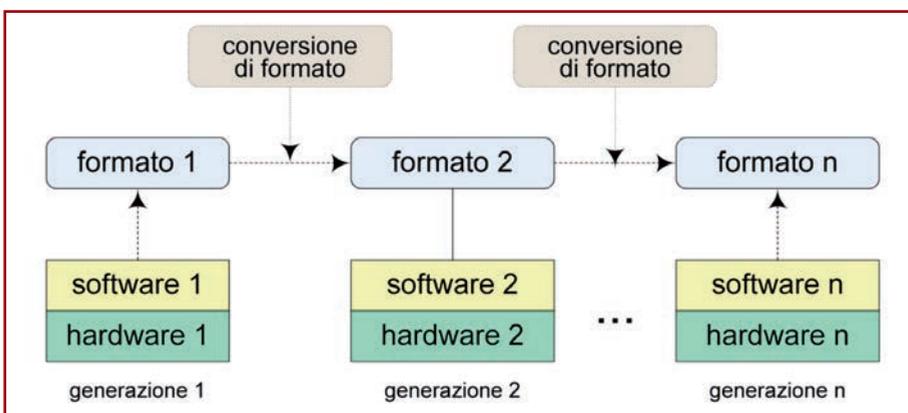


Figura 1. Schema di funzionamento della conversione di formato

¹ Cfr. Digital Preservation Coalition, *Digital Preservation Handbook*, in particolare il capitolo *File formats and standards*, <<https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>>.

² Sul significato di "proprietà significative" si veda il Paragrafo 5.

vamente frequenti, dal momento che pochi formati elettronici sono oggi in grado di “sopravvivere” più di 15-20 anni. Tuttavia, occorre tener presente che ad ogni riversamento corrisponde una perdita di informazioni – seppur minima in taluni casi –; pertanto, è fondamentale ridurre il numero dei riversamenti e indirizzare la scelta verso quei formati che si prevede siano i più “longevi” nel tempo.

Il problema dell’obsolescenza dei formati elettronici appare oggi meno grave di quanto potesse apparire alcuni decenni fa, grazie anche a un’augmentata attenzione nei confronti di questa criticità, ma non deve comunque essere sottovalutato. È importante saper individuare con precisione il momento in cui è necessario convertire i formati elettronici, scegliere i formati più idonei per la conversione, utilizzare le corrette metodologie dal punto di vista operativo. Il presente contributo intende affrontare questo problema evidenziando le motivazioni alla base della conversione di formato (perché), il momento in cui tale conversione deve essere effettuata (quando) e le metodologie per attuarla (come), facendo riferimento sia alle norme e alle linee guida internazionali, sia alle linee guida italiane, come le *Linee guida sulla formazione, gestione e conservazione dei documenti informatici* pubblicate dall’Agenzia per l’Italia Digitale (in particolare con riferimento all’Allegato 2, *Formati di file e riversamento*)³, fino a proporre una metodologia per effettuare la conversione massiva dei formati elettronici basata sulla certificazione di processo.

2. Alcune precisazioni terminologiche

Prima di proseguire nella trattazione, è necessario fare alcune precisazioni terminologiche. Nella legislazione italiana, l’operazione di conversione del formato elettronico è stata formalmente definita con la Deliberazione del Centro nazionale per l’informatica nella pubblica amministrazione (CNIPA) n. 11 del 19 febbraio 2004⁴ che introduceva i concetti di “riversamento diretto” e “riversamento sostitutivo”. Infatti, tale Deliberazione definiva il “riversamento diretto” come il «processo che trasferisce uno o più documenti conservati da un supporto ottico di memorizzazione ad un altro, non alterando la loro rappresentazione informatica»⁵ (ovvero il formato elettronico), mentre il “riversamento sostitutivo” era definito come il «processo che trasferisce uno o più documenti conservati da un supporto ottico di memorizzazione ad un altro, modificando la loro rappresentazione informatica»⁶. Mentre

³ Cfr. Agenzia per l’Italia Digitale (AgID), *Linee guida sulla formazione, gestione e conservazione di documenti informatici*, <https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_sul_documento_informatico.pdf>, in particolare l’Allegato 2, *Formati di file e riversamento*, <https://www.agid.gov.it/sites/default/files/repository_files/allegato_2_formati_di_file_e_riversamento.pdf>.

⁴ Deliberazione CNIPA n. 11/2004 del 19 febbraio 2004, *Regole tecniche per la riproduzione e conservazione di documenti su supporto ottico idoneo a garantire la conformità dei documenti agli originali*, <https://www.agid.gov.it/sites/default/files/repository_files/circolari/deliberazione-cnipa-19-02-04_0.pdf>.

⁵ Ivi, p. 2.

⁶ *Ibidem*.

per il primo processo non erano stabilite particolari modalità operative, per il secondo erano previste le modalità descritte nell'art. 3, comma 2⁷, della Deliberazione in questione, che riguardava la conservazione sostitutiva di documenti informatici. Nel 2021, con la pubblicazione delle *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici* da parte dell'Agenzia per l'Italia Digitale (AgID)⁸, sono stati introdotti i concetti di "riversamento di formato" e "riversamento di storage"; inoltre, i due concetti sono stati "unificati" introducendo la definizione di "riversamento di file" che comprende entrambe le operazioni (e, come minimo, il riversamento sostitutivo). Infatti, l'Allegato 2 precisa che «il riversamento di formato [...] comporta il trasferimento di un documento informatico in un formato di file diverso: contestualmente, ciò può comportare anche una duplicazione da un sistema di storage ad un altro (volume, disco, nastro, filesystem, storage ad oggetti o altri). Quando si parla di riversamento di file, s'intende un riversamento di almeno il suo formato»⁹. L'Allegato 2 specifica anche che «l'obsolescenza dei formati di file e dei dispositivi di archiviazione può essere mitigata soltanto mediante riversamenti periodici da una tecnologia verso un'altra; questo vale tanto per le tecnologie di storage che per i formati dei file»¹⁰. Inoltre, «le strategie relative alla gestione dei formati e dei sistemi di storage vanno spesso di pari passo; perciò, può essere in alcuni casi operativamente ed economicamente vantaggioso effettuare i riversamenti di formato contestualmente a quelli di storage. I riversamenti vengono pianificati a seguito di una nuova valutazione di interoperabilità, considerando i costi e i benefici di una tale operazione»¹¹.

Purtroppo, la terminologia utilizzata in letteratura, soprattutto quella anglosassone, non risulta sempre omogenea. Ad esempio, lo standard ISO 13008¹², sia nella prima versione, pubblicata nel 2012, che nella successiva, rilasciata nel 2022, utilizza i termini di "conversione" e "migrazione". La prima (*conversion*) è definita come «the process of changing records from one format to another»¹³, mentre la seconda (*migration*) è definita come «the process of moving records from one

⁷ Ivi, art. 3 (*Conservazione sostitutiva di documenti informatici*), comma 2: «Il processo di riversamento sostitutivo di documenti informatici conservati avviene mediante memorizzazione su altro supporto ottico e termina con l'apposizione sull'insieme dei documenti o su una evidenza informatica contenente una o più impronte dei documenti o di insiemi di essi del riferimento temporale e della firma digitale da parte del responsabile della conservazione che attesta il corretto svolgimento del processo. Qualora il processo riguardi documenti informatici sottoscritti [...] è inoltre richiesta l'apposizione del riferimento temporale e della firma digitale, da parte di un pubblico ufficiale, per attestare la conformità di quanto riversato al documento d'origine».

⁸ Le *Linee guida* sono entrate in vigore il 1° gennaio 2022.

⁹ Cfr. Allegato 2, par. 3.3 (*Riversamento*), punto 1, p. 150.

¹⁰ Cfr. Allegato 2, cap. 3 (*Raccomandazioni sui formati di file*), p. 146.

¹¹ Cfr. Allegato 2, par. 3.3 (*Riversamento*), p. 150.

¹² Cfr. ISO 13008:2022 – *Information and documentation – Digital records conversion and migration processes*.

¹³ Cfr. ISO 13008:2022, par. 4.2 – *Conversion and migration drivers*, p. 4.

hardware or software configuration to another without changing the format»¹⁴. In sostanza, la prima operazione consiste nella conversione di un documento da un formato elettronico ad un altro e corrisponde al riversamento sostitutivo della citata Deliberazione CNIPA; la seconda operazione consiste nel trasferimento di un documento da un supporto di archiviazione ad un altro senza alterarne il formato elettronico e corrisponde al riversamento diretto della Deliberazione.

Analoghe definizioni vengono fornite dalla norma ISO 15489-1¹⁵ – secondo la quale “conversion” è «[the] process of changing records from one format to another» e “migration” è «[the] process of moving records from one hardware or software configuration to another without changing the format»¹⁶ – così come dalla norma ISO 30300¹⁷ secondo la quale “conversion” significa «changing records from one format to another» mentre “migration” consiste in «moving records from one hardware or software configuration to another»¹⁸. Dello stesso avviso è il dizionario prodotto nell’ambito del progetto InterPARES 2¹⁹, il quale definisce “conversion” come «the process of transforming a digital document or other digital object from one format, or format version, to another one» e “migration” come «the process of moving or transferring digital objects from one system to another».

Al contrario, il dizionario di terminologia archivistica della Society of American Archivists (SAA) definisce “format migration” come «the practice of converting an electronic file to a different standard file type to circumvent obsolescence», ovvero l’operazione che le fonti precedentemente citate chiamano “conversion”. Una nota alla voce precisa che «in format migration, the content is preserved but the bits are not. In past use, the term has been less precise and could have included media migration». Lo stesso termine, “migration”, è utilizzato anche nel caso di trasferimento da un supporto di archiviazione ad un altro; infatti, il dizionario definisce “media migration” come «the practice of copying records from one physical carrier to another for preservation» ma attribuisce sostanzialmente lo stesso significato anche al termine “conversion”, definito come il processo «to move data to a different format, especially data from an obsolete format to a current format». Il dizionario prevede una voce anche per il termine “migration” che è definito come

¹⁴ *Ibidem*.

¹⁵ Cfr. ISO 15489-1:2016 – *Information and documentation – Records management – Part 1: Concepts and principles*.

¹⁶ Si noti che nell’edizione 2001 della norma ISO 15489 i concetti non erano così ben definiti: infatti, l’operazione di “conversion” era definita come «the process of changing records from one medium to another or from one format to another», mentre l’operazione di “migration” era definita come «the act of moving records from one system to another, while maintaining the records’ authenticity, integrity, reliability and useability».

¹⁷ Cfr. ISO 30300:2022, *Information and documentation – Records management – Core concepts and vocabulary*, par. 3.4 *Terms relating to the concepts of records management and records processes*, p. 8.

¹⁸ *Ibidem*.

¹⁹ Disponibile all’indirizzo <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf>.

«the process of moving data from one information system or storage medium to another to ensure continued access to the information as the system or medium becomes obsolete or degrades over time».

Lo standard ISO 14721²⁰ tratta in maniera approfondita il tema del riversamento, a cui dedica l'intero Capitolo 5 (*Preservation Perspectives*) nel quale vengono descritte le diverse tipologie di operazioni di riversamento. La terminologia utilizzata è quella di "digital migration", definita come «the transfer of digital information, while intending to preserve it, within the OAIS»²¹. Il modello OAIS prevede complessivamente quattro tipologie di operazioni di «digital migration» dei pacchetti informativi: le prime due non producono cambiamenti nella sequenza di bit, mentre le ultime due la modificano:

- l'operazione di "refreshment" che consiste nella sostituzione di un supporto di archiviazione con un altro dello stesso tipo e nel trasferimento degli AIP²² dall'uno all'altro senza alterarne la sequenza;
- l'operazione di "replication", che duplica uno o più AIP sullo stesso o su altro supporto di archiviazione, senza alterarne le informazioni di impacchettamento, il contenuto informativo e le informazioni di conservazione; si noti che l'operazione di refreshment di fatto è un'operazione di replication, ma quest'ultima può richiedere modifiche alla mappatura del sistema di archiviazione;
- l'operazione di "repackaging", che comporta la modifica delle informazioni di impacchettamento relative ad un AIP;
- l'operazione di "transformation", che determina la variazione del contenuto informativo e/o delle informazioni di conservazione relative a un AIP; La conversione del formato elettronico è un tipico esempio di operazione di "transformation" e si tratta di un'operazione che può essere reversibile o non reversibile;

Anche il *Digital Preservation Handbook*, pubblicato dalla Digital Preservation Coalition (DPC), utilizza solo il termine "migration" e distingue tra "format migration" e "storage media migration"; il primo è definito come «transferring or transforming (i.e. migrating) data from an ageing/obsolete format to a new format, possibly using new applications systems at each stage to interpret information» e per

²⁰ ISO 14721:2012 – *Space data and information transfer systems – Open archival information system (OAIS) – Reference model* è la norma che recepisce il modello concettuale OAIS che è oggi alla base di qualunque sistema di conservazione digitale. L'ultima versione dello standard è stata pubblicata nel 2012.

²¹ Cfr. ISO 14721:2012, par. 5.2 (*Digital migration*), p. 101.

²² Gli AIP (Archival Information Package) sono i pacchetti di archiviazione, una delle tipologie di pacchetti informativi impiegati dal modello OAIS. Secondo tale modello, il pacchetto di archiviazione è un derivato del pacchetto di versamento (SIP, Submission Information Package) o di un insieme di pacchetti di versamento che vengono affidati all'OAIS con il fine di archiviare i dati contenuti. Per il significato dei termini qui utilizzati si rinvia allo standard ISO 14721. Per una trattazione approfondita del modello OAIS si veda Stefano Pigiapoco, *Progetto Archivio digitale, Metodologia Sistemi Professionalità*, Lucca: Civita editoriale, 2018.

chiarire meglio il significato viene anche riportato un esempio: «moving from one version of a format standard to a later standard is a version of this method; for example moving from MS Word Version 6 (from 1993) to MS Word for Windows 2010»²³. Il secondo è chiaramente riferito al riversamento da un supporto di archiviazione ad un altro.

In sintesi, alcuni autori utilizzano solo il termine “migration”, applicandolo di volta in volta alla migrazione da un formato elettronico ad un altro (laddove, a parere dello scrivente, sarebbe più corretto utilizzare il termine “conversion”) oppure da un supporto di archiviazione ad un altro. Non a caso Fleischhauer e Bradley hanno scritto che:

«In the field of digital preservation, the term migration is used in two ways. Media or system migration refers to the movement of digital files from obsolete data-storage media or an obsolete data-management system to new media or a new system. Media migration is sometimes called physical migration and media refreshment. In this form of migration “the bits do not change”. In contrast, format migration, also known as logical migration, refers to the movement of a content item from one format to another: “the bits do change”»²⁴.

La situazione è comunque molto variegata: infatti, vi sono anche autori che utilizzano i termini “migration” e “conversion” in maniera intercambiabile, altri che, invece, attribuiscono loro significati opposti. Tutto questo, purtroppo, contribuisce a ingenerare un certo livello di confusione in un settore che, invece, è già piuttosto complesso e avrebbe bisogno di molta più chiarezza.

Di recente, anche il *Regolamento (UE) 2024/1183 del Parlamento europeo e del Consiglio dell'11 aprile 2024*²⁵ ha fatto riferimento in diversi punti ai concetti di riversamento diretto e sostitutivo, come operazioni necessarie per assicurare la *long-term preservation*. In particolare, viene specificato che

«è opportuno istituire un quadro giuridico per i servizi di archiviazione elettronica qualificati, ispirato al quadro per gli altri servizi fiduciari di cui al presente regolamento. [...] Ove necessario, tali disposizioni dovrebbero consentire che i dati elettronici e i documenti elettronici conservati siano trasferiti su supporti o formati diversi al fine di estenderne la durabilità e la leggibilità oltre il periodo di validità tecnologica, evitando nel contempo, nella misura del possibile, le perdite e le alterazioni [...]»²⁶.

²³ Cfr. Digital Preservation Coalition, *Digital Preservation Handbook*, in particolare il capitolo *Preservation action*, <<https://www.dpconline.org/handbook/organisational-activities/preservation-action>>.

²⁴ Cfr. *Guidelines for the Preservation of Video Recordings, IASA-TC 06*, a cura di C. Fleischhauer e K. Bradley, London (UK): International Association of Sound and Audiovisual Archives, 2019, <https://www.iasa-web.org/sites/default/files/publications/IASA-TC_06-A_v2019.pdf>.

²⁵ Il *Regolamento (UE) 2024/1183 del Parlamento europeo e del Consiglio dell'11 aprile 2024 che modifica il regolamento (UE) n. 910/2014 per quanto riguarda l'istituzione del quadro europeo relativo a un'identità digitale*, è stato pubblicato nella Gazzetta Ufficiale della Comunità Europea del 30 aprile 2024. È conosciuto anche come “Regolamento eIDAS 2”.

²⁶ Cfr. *Regolamento (UE) 2024/1183*, Considerando 66.

Le operazioni di riversamento diretto e sostitutivo vengono ricordate anche all'articolo 45 undecies laddove, tra i requisiti che i servizi di archiviazione elettronica qualificati devono soddisfare, viene elencato quello di assicurare che i dati elettronici e i documenti elettronici «siano conservati in modo tale da essere protetti dal rischio di perdita e alterazione, ad eccezione delle modifiche riguardanti il loro supporto o il loro formato elettronico»²⁷.

Nel seguito, verrà adottata la terminologia proposta dalla norma ISO 13008 e quindi si utilizzeranno le locuzioni “conversione di formato” e “migrazione di supporto”; a tal proposito è bene ricordare che nel contesto italiano la terminologia prevista dall'attuale normativa è quella di “riversamento di formato” e “riversamento di storage”. Inoltre, si tratterà, per semplicità, di “documenti digitali” (spesso abbreviati in “documenti” per non appesantire il discorso) e di “archivi digitali” ma resta inteso che i concetti e le riflessioni che verranno esposti si possono applicare, in generale, a tutte le tipologie di oggetti digitali, tanto a quelli presenti in un archivio digitale quanto a quelli presenti in una biblioteca digitale o in qualsiasi altro repository.

3. Perché convertire i formati elettronici

Quando si parla di conversione di formato elettronico, il primo aspetto da prendere in considerazione è il motivo per cui si rende necessario procedere alla conversione. Secondo lo standard ISO 13008, le ragioni sono riconducibili a quattro fattispecie²⁸:

- a) obsolescenza: i documenti sono codificati secondo formati obsoleti ma ancora leggibili e quindi devono essere convertiti in formati più moderni²⁹;
- b) questioni di proprietà: i documenti sono codificati secondo formati elettronici proprietari e devono quindi essere convertiti in formati non proprietari³⁰;
- c) motivi di interoperabilità: i documenti devono essere convertiti in un formato elettronico che garantisca la perfetta interoperabilità con determinate infrastrutture tecnologiche;
- d) motivi legali: i documenti devono essere convertiti in base a espliciti requisiti legali o normativi riguardanti i formati elettronici o i fornitori di servizi.

Quando si converte il formato elettronico di un documento, il risultato finale può essere la *sostituzione* del documento originario con un altro documento in un for-

²⁷ Cfr. *Regolamento (UE) 2024/1183*, art. 45 undecies (*Requisiti per i servizi di archiviazione elettronica qualificati*).

²⁸ Cfr. ISO 13008, par. 4.2.2 (*Conversion drivers*), p. 4.

²⁹ Ad esempio, potrebbe essere necessario convertire gli oggetti codificati in un formato di immagine obsoleto (MSP, PCX, TGA ecc.) in un formato elettronico più aggiornato (JPG, WEBP, HEIC ecc.).

³⁰ Come nel caso della conversione di documenti testuali nel formato DOC (il “vecchio” formato Microsoft per i documenti di testo) in DOCX o PDF/A.

mato diverso oppure la *creazione* di una versione aggiuntiva del documento in un formato diverso per soddisfare i requisiti di usabilità³¹.

Il primo scenario si verifica quando la conversione è necessaria per mantenere l'accesso ai documenti dell'archivio digitale e assicurarsi che siano pienamente disponibili e utilizzabili nel tempo. Ad esempio, ciò può capitare a seguito di cambiamenti negli strumenti software utilizzati nell'archivio digitale, dell'abbandono di formati legacy e a rischio di obsolescenza o di cambiamenti nel formato standard utilizzato dall'archivio digitale per la pubblicazione online. A volte è necessario convertire i formati elettronici non solo perché diventano obsoleti, ma anche per ragioni legate ai cambiamenti tecnologici. Ad esempio, un formato elettronico potrebbe essere ancora attuale ma è necessario convertirlo perché l'ambiente tecnologico utilizzato per gestire l'archivio digitale è cambiato. Se i formati dei file non venissero convertiti con un approccio proattivo, nel tempo ci si potrebbe trovare nell'impossibilità di accedere ai documenti o di utilizzarli nel modo desiderato, oppure nella necessità di utilizzare software particolari e costosi per poterne recuperare l'accesso. Quando si convertono i formati elettronici, occorre fare attenzione al fatto che i documenti originali potrebbero essere eliminati, e ciò potrebbe comportare dei rischi nel lungo periodo.

Nel secondo scenario, anziché convertire un documento in un nuovo formato, vengono create versioni aggiuntive di quel documento in formati diversi per consentire nuove forme di accesso e di utilizzo, come la condivisione o la pubblicazione di informazioni, l'utilizzo delle informazioni contenute nei documenti in modi nuovi, l'aggregazione di informazioni provenienti da varie fonti. Ciò non implica che il formato originale sia obsoleto ma solo che è necessario avere lo stesso documento in più di un formato per soddisfare determinati requisiti. Ad esempio, ciò si verifica quando un documento creato utilizzando un formato di elaborazione testi (come il DOCX) deve essere convertito in un altro formato (come il PDF o il PDF/A) per essere pubblicato online. Oppure quando, in un progetto di digitalizzazione di materiali cartacei, vengono acquisite delle immagini in formato TIFF alla più alta qualità possibile per finalità di conservazione (formato "master") e da queste vengono prodotte – mediante il processo di conversione – delle immagini in formato JPG, eventualmente con risoluzioni e qualità diverse, per finalità di fruizione (formati "derivati") oppure delle immagini in formato GIF da utilizzare come miniature. Tuttavia, occorre tener presente che i formati non dovrebbero essere moltiplicati inutilmente: se un unico formato soddisfa tutte le esigenze di accesso, di solito è la soluzione migliore³².

³¹ Cfr. The National Archives of United Kingdom, *File Format Conversion*, p. 7 e p. 11, <<https://cdn.nationalarchives.gov.uk/documents/information-management/format-conversion.pdf>>.

³² Ivi, p. 11.

4. Quando convertire i formati elettronici

Una volta stabilito che è necessario convertire i documenti da un formato elettronico ad un altro, occorre decidere quando è opportuno effettuare questa conversione. Vi sono fondamentalmente tre metodologie, che dipendono in gran parte dalle motivazioni che sono alla base del processo di conversione (ma che possono anche dipendere dall'ambiente tecnico o da altre esigenze del sistema di conservazione)³³: 1) la conversione su richiesta; 2) la conversione anticipata; 3) la conversione tardiva. È importante analizzarle nel dettaglio:

1) *Conversione su richiesta (on demand)*. Con questa strategia la conversione è di tipo dinamico in quanto viene effettuata "al bisogno", ogni volta che viene ricevuta una richiesta in tal senso. In genere si opera su un singolo documento alla volta, ma è applicabile anche alle conversioni di tipo massivo (batch). Questa strategia può essere utilizzata anche per sostituire dei documenti con altri in formati diversi, ma più spesso è utilizzata per creare versioni aggiuntive dei documenti, a seconda delle necessità³⁴. I vantaggi sono molteplici: è necessario archiviare un solo documento e non le diverse versioni di quel documento nei vari formati richiesti, quindi si minimizza la necessità di spazio di archiviazione; non è necessario convertire un gran numero di documenti tutti in una sola volta, operazione che potrebbe essere onerosa per il sistema e richiedere molto tempo; l'aggiunta di nuovi documenti al sistema è semplice, in quanto non è necessario fornire in anticipo tutti i formati richiesti; il sistema può essere aggiornato per fornire formati diversi in base alle esigenze, anche in questo caso senza dover elaborare in anticipo tutti i documenti esistenti. Tuttavia, ci sono anche degli aspetti negativi: non c'è quasi modo di verificare la qualità dei file convertiti: se si adotta questa strategia, è necessario assicurarsi che il processo di conversione sia sufficientemente affidabile per le proprie esigenze; la conversione on-demand può essere lenta o eccessivamente onerosa per i sistemi, a seconda delle dimensioni, della complessità e del numero di conversioni, costringendo l'utente a delle attese eccessive.

2) *Conversione anticipata*. Con questa strategia la conversione dei documenti in formati diversi viene eseguita non appena ciò è opportuno (ma non su richiesta). Essa viene effettuata non per finalità di conservazione ma per finalità di gestione, in quanto l'obiettivo è quello di convertire i documenti in un formato che meglio si adatta alle esigenze dell'archivio digitale, anche se il formato precedente non è

³³ Ivi, p. 14.

³⁴ Ad esempio, questo tipo di strategia viene applicata quando è necessario offrire agli utenti lo stesso documento ma in diversi formati (come DOCX, ODT, RTF). In molti casi, non è conveniente archiviare ogni documento in tutti i formati, ma è meglio archivarlo in un solo formato (di solito nel formato più completo) e generare gli altri quando si riceve una richiesta di fruizione di quel documento in un formato diverso.

ancora diventato obsoleto. Ad esempio, se il responsabile dell'archivio ha deciso di utilizzare un nuovo formato perché è quello richiesto da un software aggiornato, è possibile convertire tutti i documenti dal precedente formato al nuovo. Questa strategia presenta molti vantaggi. Il numero dei differenti formati elettronici da gestire si riduce notevolmente convertendo i documenti in uno o più formati standardizzati³⁵; si riducono i costi di supporto, manutenzione e licenza software; il rischio di obsolescenza del formato diventa trascurabile; è possibile verificare con accuratezza che il processo di conversione si sia svolto nel modo corretto e garantire la qualità. Ovviamente ci sono anche degli svantaggi: i documenti sono convertiti più frequentemente e, aumentando il numero di conversioni, aumentano parallelamente sia il rischio di perdita di informazioni che i costi associati alle operazioni di riversamento³⁶; se i nuovi formati sono piuttosto recenti, gli strumenti di conversione potrebbero non essere così facilmente disponibili, potrebbero avere dei bug o non essere in grado di gestire bene documenti particolarmente complessi, ed anche questo aspetto può influire sul costo e sulla qualità del processo di conversione; il nuovo formato potrebbe non essere così ampiamente adottato, per cui potrebbe rendersi necessaria la produzione di ulteriori formati per facilitare la condividere dei documenti con utenti che non hanno ancora effettuato l'aggiornamento³⁷.

3) *Conversione tardiva*. Con questa strategia la conversione è rimandata fino all'ultimo momento utile. Naturalmente, la definizione di "ultimo momento utile" varia notevolmente a seconda della valutazione dei rischi e dei benefici dell'archivio digitale. Questa strategia presenta molti vantaggi: i documenti vengono convertiti meno frequentemente minimizzando così il rischio di perdita di informazioni e riducendo i costi associati all'operazione di conversione; se il formato di destinazione è ampiamente adottato, sarà probabilmente disponibile un maggior numero di strumenti di conversione e questi saranno verosimilmente in grado di gestire meglio i documenti insoliti o complessi, perché c'è stato il tempo di risolvere bug e casi par-

³⁵ Ad esempio, questo tipo di strategia viene applicata quando si decide di convertire tutti i documenti testuali nei formati DOCX o ODT (che non sono certamente obsoleti e sono considerati compatibili con un processo di conservazione a lungo termine dal già citato *Allegato 2 alle Linee Guida sulla formazione, gestione e conservazione dei documenti informatici* dell'AgID) nell'unico formato PDF/A.

³⁶ «While format obsolescence is a real risk to digital content, practitioners should not be overly zealous with format migrations, particularly for files in formats not at high risk. Institutions should define local file format policies based on calculated risk assessments rather than relying on the default file format policies implemented in digital preservation software systems». Cfr. Keith L. Pendergrass et al., *Toward Environmentally Sustainable Digital Preservation*, «American Archivist», 82 (Spring/Summer 2019), n. 1, p. 189, <<https://doi.org/10.17723/0360-9081-82.1.165>>.

³⁷ Con la conseguenza che l'archiviazione di tutti i documenti convertiti richiederà più spazio rispetto alla conversione on-demand. Cfr. The National Archives of United Kingdom, *File Format Conversion*, p. 16.

ticolari; alcuni documenti potrebbero aver superato il tempo di conservazione e quindi essere destinati allo scarto, evitando così la necessità di convertirli. Tuttavia, ci sono anche degli svantaggi: nell'archivio digitale ci sarà una maggiore varietà di formati diversi in uso e ciò può aumentare i costi di supporto, manutenzione e licenza del software, ridurre la flessibilità nella scelta di software diversi e impedire che i documenti nei formati più vecchi siano utilizzabili in contesti più recenti; potrebbe essere necessario convertire un numero maggiore di formati elettronici in contemporanea, e ciò complica la gestione della conversione e l'assicurazione della qualità; se è necessario che gli stessi documenti siano accessibili in più formati, con questa strategia l'archiviazione di tutti i documenti nei vari formati richiederà più spazio rispetto alla conversione on-demand; infine, l'individuazione dell'“ultimo momento utile” potrebbe essere non essere così precisa e si potrebbe scoprire che ormai è troppo tardi e la conversione di alcuni documenti non è più economicamente o tecnicamente fattibile.

Come si è potuto vedere, le diverse strategie (conversione on-demand, anticipata e tardiva) hanno ciascuna dei vantaggi e degli svantaggi, con rischi diversi dovuti alla tempistica della conversione. Non esiste una strategia uguale per tutti e solo valutando le esigenze dell'archivio digitale è possibile determinare il giusto equilibrio tra costi e benefici e mettere in atto la strategia più opportuna.

5. Come convertire i formati elettronici

Secondo lo standard ISO 13008, il processo di conversione di formato elettronico avviene seguendo quattro fasi fondamentali: a) pianificazione, b) verifica, c) conversione, d) convalida³⁸ (Fig. 2). Nel seguito si prenderanno in esame le caratteristiche di ciascuna fase.

a) Pianificazione (*planning*). È la fase più importante, perché le maggiori probabilità di successo del processo di conversione derivano da una pianificazione attenta e accurata. In primo luogo, è necessario valutare i documenti da convertire per comprendere quali sono le proprietà e le funzionalità che devono essere mantenute nel processo di conversione. Questi requisiti potrebbero non essere immediatamente evidenti, ed è opportuno lavorare con il responsabile dell'archivio digitale e con gli utenti per capire se i documenti nei formati da cui si sta effettuando la migrazione hanno caratteristiche particolari che si vuole garantire rimangano invariate e assicurare che tutti i loro requisiti siano soddisfatti. Alcuni processi di conversione modificano solo il formato elettronico dei documenti, ma altri ne alterano anche alcune delle proprietà. Come regola generale, si può assumere che i documenti che hanno una struttura molto semplice (come quelli in formato TXT) posso-

³⁸ Cfr. ISO 13008:2022, cap. 7 (*Conversion and migration procedures*), p. 14.

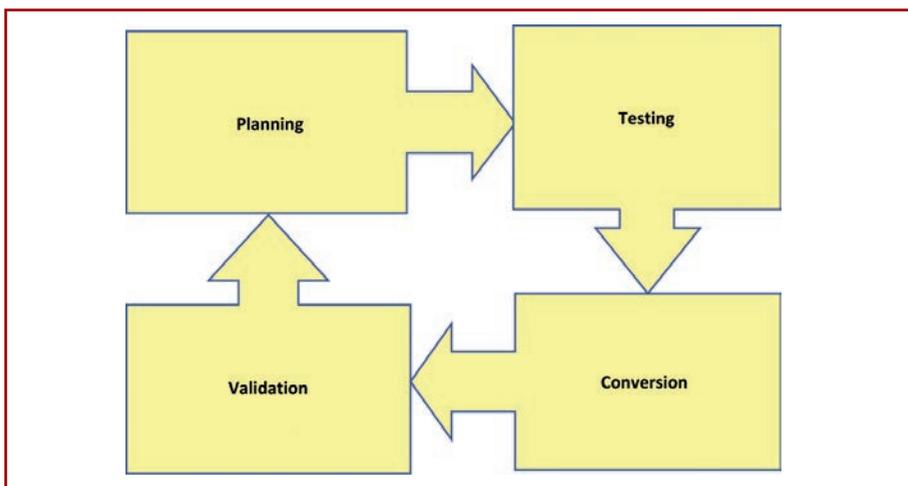


Figura 2. Le fasi del processo di conversione di formato

no superare un processo di conversione senza subire modifiche sostanziali, ma i documenti complessi subiscono quasi sempre qualche forma di alterazione³⁹.

È importante tenere presente che qualsiasi processo di conversione espone potenzialmente i documenti al rischio di perdita di informazioni. Pertanto, prima della conversione, è necessario identificare le “proprietà significative” che devono essere mantenute e devono “sopravvivere” alla conversione senza (o con poche) modifiche. A questo proposito, è importante ricordare che il progetto di ricerca che più di ogni altro si è dedicato a questi aspetti è stato InSPECT (Investigating the Significant Properties of Electronic Content Over Time), finanziato nel Regno Unito dal JISC tra marzo 2007 e marzo 2009 nell’ambito del programma “Repositories and Preservation”⁴⁰. Il progetto ha avuto come obiettivo la definizione di una meto-

³⁹ Ad esempio, nella conversione da un documento di testo ad un altro (come nel caso della conversione da un documento in formato DOCX ad un documento nel formato ODT) è possibile che il testo rimanga invariato mentre il layout, i colori, gli stili e i caratteri vengano alterati. Nel passato questo tipo di criticità di presentava abbastanza spesso, mentre oggi la differenza è solitamente trascurabile. Tuttavia, si possono verificare delle situazioni particolari, come nel caso in cui un cambiamento nell’impaginazione di un documento testuale (si pensi ad un contratto) causi delle modifiche nei riferimenti di pagina incorporati nel documento stesso rendendolo inutilizzabile. Oppure, nel caso in cui il processo di conversione alteri i colori, è possibile che si perdano le annotazioni inserite (ad esempio, il colore verde per indicare le attività completate e il colore rosso per quelle non ancora completate) e con esse delle informazioni importanti. Quindi è importante valutare con molta attenzione le informazioni del documento che devono essere mantenute. Cfr. The National Archives of United Kingdom, *File Format Conversion*, par. *Assess your information*, p. 18-20.

⁴⁰ Il progetto è stato guidato dall’Arts and Humanities Data Service (AHDS) in collaborazione con The National Archives. Successivamente, dopo la soppressione dell’AHDS nel marzo 2008, è stato guidato dal Centre for e-Research (CeRch) del Kings College di Londra, sempre in collaborazione con The National Archives. Buona parte del materiale prodotto nell’ambito del progetto è ancora oggi disponibile sul sito <<https://significantproperties.kdl.kcl.ac.uk>>.

dologia per identificare le “proprietà significative” delle varie categorie di documenti digitali. Per “proprietà significative” i promotori del progetto intendono le caratteristiche dei documenti digitali che devono essere conservate nel tempo, anche a seguito di operazioni di conversione. Ad esempio, sono proprietà significative il contenuto dei documenti, i metadati, il loro aspetto (ad esempio, il layout, i colori ecc.), lo scopo per cui sono stati prodotti o anche la loro struttura logica⁴¹. Queste proprietà significative variano a seconda della categoria di documento digitale⁴²: il progetto si è concentrato su quattro categorie di documenti: 1) le immagini raster; 2) le e-mail; 3) i documenti di testo; 4) i documenti sonori.

È importante assicurarsi che i formati elettronici verso i quali convertire i documenti siano in grado di supportare le “proprietà significative” e che il processo di conversione sia in grado di mantenerle durante la conversione. Se il nuovo formato non supportasse le caratteristiche richieste, si renderebbe necessario rivalutare la scelta del formato elettronico. Spesso si devono considerare anche alcune caratteristiche meno ovvie, di solito legate a funzionalità complesse o nascoste del formato. Di seguito è riportato un elenco non esaustivo di alcune di esse⁴³:

– *metadati incorporati*. Molti formati consentono di incorporare vari metadati descrittivi. Ad esempio, alcuni formati per documenti testuali incorporano dei metadati per identificare l’autore, la data di creazione, la data di ultima modifica ecc.; alcuni formati per fotografie incorporano dei metadati che indicano le impostazioni della fotocamera, la posizione geografica al momento dello scatto ecc.⁴⁴ Spesso si tratta di informazioni rilevanti ed è opportuno valutare se sia necessario convertire anche tali metadati incorporati e se gli strumenti di conversione siano in grado di trasferirli;

– *oggetti incorporati*. Molti formati complessi consentono di incorporare degli oggetti digitali in vari formati. Ad esempio, i documenti di testo possono contenere immagini o fogli di calcolo incorporati; le presentazioni possono contenere contenuti audio e video. Non tutti gli strumenti di conversione sono in grado di gestire correttamente tutti i tipi di oggetti incorporati. Pertanto, è necessario effettuare delle verifiche su documenti con oggetti incorporati per garantire la qualità del processo di conversione;

⁴¹ Cfr. la scheda relativa al progetto disponibile sul sito del Polo archivistico dell’Emilia-Romagna (ParER), <<https://poloarchivistico.regione.emilia-romagna.it/saperi/progetti/inspect-investigating-the-significant-properties-of-electronic-content-over-time>>.

⁴² Ad esempio, nei formati per documenti testuali sarà importante considerare il contenuto, il tipo di carattere ecc.; nei formati per immagini sarà importante la risoluzione, la profondità di colore ecc.; nei formati per documenti sonori sarà importante la frequenza di campionamento, la profondità di suono e così via.

⁴³ Cfr. The National Archives of United Kingdom, *File Format Conversion*, par. *Assess your information*, p. 18-20.

⁴⁴ Questo è il caso dei metadati EXIF nel caso delle fotografie in formato JPG o dei metadati ID3 nel caso dei documenti sonori in formato MP3.

- *script e macro*. Alcuni formati possono contenere codice scritto con dei linguaggi di programmazione. Ad esempio, i documenti di *office automation* possono includere delle macro per automatizzare le attività più comuni. In generale, gli script e le macro non “sopravvivono” ai processi di conversione, a meno che non si tratti di conversioni da una versione di un formato a un’altra versione dello stesso formato⁴⁵. Pertanto, se il supporto di script e macro è imprescindibile, potrebbe essere necessario riscriverli manualmente nella versione prevista dal nuovo formato;
- *firme digitali*. Alcune categorie di documenti consentono di incorporare firme digitali al loro interno. Nella conversione di formato, cambia la sequenza binaria; pertanto, dopo la conversione la firma digitale perderà la sua validità e sarà necessario individuare delle strategie per mantenere il valore giuridico.

Un passo molto importante nella fase di pianificazione è la scelta dello strumento di conversione dei documenti; si tratta di una scelta difficile, perché ne esistono numerosi, sia proprietari, che freeware o open-source. Tuttavia, non per tutti le categorie di formati elettronici vi è lo stesso livello di “copertura”. Per i formati più diffusi, come le immagini, sono disponibili diversi programmi, ma per i formati di nicchia o per quelli più vecchi le scelte possono essere molto limitate. Per i formati con scarso supporto potrebbe essere necessario eseguire due conversioni, utilizzando un formato elettronico intermedio per colmare il divario tra il formato iniziale e quello desiderato⁴⁶. In alcuni casi, potrebbe essere necessario commissionare un software ad hoc per eseguire la conversione, soprattutto se i formati elettronici sono molto specifici⁴⁷. È importante valutare se il software è in grado di convertire pienamente le proprietà significative del documento e i relativi metadati, e non solo se “semplicemente” converte dal formato di origine a quello di destinazione.

b) Verifica (*testing*). Una volta individuati gli strumenti da utilizzare e prima di avviare il processo di conversione nella sua interezza, è bene effettuare una fase di test su un campione rappresentativo dei documenti per verificare che le loro proprietà significative e i metadati vengano accuratamente convertiti e senza perdita di autenticità, affidabilità, integrità e usabilità. Il test richiede una conoscenza accurata dei formati elettronici e della configurazione hardware o software di partenza e di quelli di destinazione. Il test deve assicurare non solo che venga effettuata con un livello di quali-

⁴⁵ Ad esempio, dal formato DOCX di Microsoft Word 2007 al formato DOCX di Microsoft Word 2013.

⁴⁶ Ad esempio, nel caso si dovessero convertire dei vecchi documenti di testo creati negli anni '80 del secolo scorso utilizzando il “famoso” programma WordStar, si potrebbe ipotizzare una prima conversione dal formato di WordStar (.WS) al formato delle prime versioni di Microsoft Word (.DOC) e poi una seconda conversione da quest’ultimo al formato introdotto con Microsoft Word 2007 (.DOCX).

⁴⁷ Lo sviluppo di un software in grado di convertire da un certo formato elettronico ad un altro è possibile solo se sono disponibili le specifiche del formato di entrambi; è per questo che è preferibile utilizzare formati “aperti”. Cfr. nota 69.

tà accettabile la conversione del documento, ma anche che i metadati vengano convertiti con lo stesso livello di qualità; a tal fine è necessario utilizzare degli strumenti di estrazione dei metadati per confrontare quelli di partenza con quelli di destinazione. Potrebbe anche essere necessario l'utilizzo di strumenti di estrazione dei metadati diversi per i formati di origine e di destinazione e convertire i loro risultati in una forma comune per facilitare il confronto. Infine, sarebbe auspicabile individuare delle metriche per misurare in maniera automatica o semi-automatica il livello di qualità della conversione e poter scegliere lo strumento software più valido⁴⁸.

c) Conversione (*conversion*). Una volta acquisita una sufficiente comprensione delle metodologie da seguire e selezionati i formati e gli strumenti, si è pronti per iniziare l'operazione di conversione dei documenti. In questa fase è molto importante impostare correttamente i parametri dello strumento software di conversione⁴⁹. Alcuni strumenti di conversione operano solamente sul singolo documento; se si desidera convertire più documenti, è necessario scrivere uno script per automatizzare l'elaborazione di un lotto di documenti oppure utilizzare degli strumenti che consentono di operare su interi lotti di documenti. Oggi esistono anche numerose aziende che eseguono la conversione dei formati elettronici come servizio, offrendo anche attraverso un'interfaccia online e assicurando la confidenzialità dei documenti inviati per la conversione.

d) Convalida (*validation*). A conclusione del processo di conversione è necessario un ulteriore processo di validazione (oltre quello del test iniziale) per verificare se la conversione è stata eseguita in maniera corretta e, in particolare, se le proprietà significative e i metadati sono rimasti effettivamente invariati o se, al contrario, si sono verificate delle perdite. Ovviamente, il controllo manuale di ogni documento può essere impraticabile; quindi, la soluzione migliore è quella di effettuare un controllo su un campione rappresentativo di documenti convertiti, confrontando il documento originale con quello convertito⁵⁰. A seconda della tipologia di docu-

⁴⁸ Tra i progetti che si sono occupati della garanzia della qualità si segnalano AQUA (Automating Quality Assurance), <<https://wiki.opf-labs.org/display/AQuA/Home>>; SPRUCE (Sustainable Preservation Using Community Engagement), <<https://wiki.opf-labs.org/display/SPR/Home>>; SCAPE (Scalable Preservation Environments), <<https://scafe-project.eu>>. Si veda anche Paul Wheatley — Bo Middleton — Jodie Double — Andrew N. Jackson — Rebecca McGuinness, *People Mashing: Agile Digital Preservation and the AQUA Project*, <<https://services.phaidra.univie.ac.at/api/object/o:294255/download>>.

⁴⁹ Cfr. Peter Bajcsy et al., *A Framework for Understanding File Format Conversions*, in: *US-DPIF '10: Proceedings of the 2010 Roadmap for Digital Preservation Interoperability Framework Workshop*, New York, (NY): Association for Computing Machinery, 2010, p. 1-7, <<https://doi.org/10.1145/2039274.2039284>>.

⁵⁰ Il confronto riguarderà in prima battuta le proprietà di base del documento originale e di quello convertito. Ad esempio, per un'immagine è possibile estrarre le dimensioni (altezza e larghezza) e verificare che siano le stesse prima e dopo la conversione; per un documento sonoro è possibile esami-

mento, questo tipo di confronto diretto può essere di varia natura. Ad esempio, mentre per le immagini è possibile effettuare un'analisi a livello visivo, per gli oggetti sonori sarà necessario ascoltare il documento originale e quello trasferito per capire se ci sono differenze. Naturalmente, si possono utilizzare strumenti di confronto automatici o addirittura basati sull'intelligenza artificiale. Allo stesso modo occorre confrontare, utilizzando degli appositi strumenti di estrazione, i metadati del documento originale e di quello convertito per verificare che anch'essi siano stati convertiti correttamente. Anche gli utenti finali dovrebbero essere coinvolti in questo processo di verifica, in quanto potrebbero rilevare dei problemi poco evidenti che altrimenti passerebbero inosservati.

Anche se i documenti originali sono destinati ad essere eliminati dopo il processo di conversione, è buona norma conservarli per un certo periodo di tempo, onde ridurre il rischio che nel tempo possano emergere proprietà significative che non erano state prese in considerazione e che quindi si sono perse nel processo di conversione. Questo vale anche se i processi di verifica della qualità hanno dimostrato che la conversione è pienamente riuscita. Non è possibile stabilire a priori per quanto tempo gli originali debbano essere conservati, poiché ciò dipenderà dall'importanza dei documenti, dalle motivazioni alla base della conversione, dalla tolleranza al rischio dell'organizzazione, dalla sua fiducia nel processo di conversione, dai costi di conservazione degli originali e di mantenimento delle relazioni tra questi e i documenti convertiti e dall'esigenza di ridurre il numero di documenti archiviati.

6. Il riversamento di formato elettronico secondo le Linee guida AgID

Nel contesto italiano, il principale riferimento sul tema della conversione dei formati elettronici è rappresentato dall'*Allegato 2 alle Linee guida sulla formazione, gestione e conservazione dei documenti informatici* pubblicate dall'Agenzia per l'Italia Digitale (AgID) e divenute obbligatorie a partire dal 1° gennaio 2022. In base a tali *Linee guida*, nel caso di documenti informatici codificati secondo formati elettronici ormai diventati obsoleti o deprecati, «possono essere necessarie attività di riversamento dei documenti in altro formato diverso da quello originale»⁵¹ al fine di assicurarne la conservazione nel tempo⁵². Tale riversamento «può avvenire

nare la lunghezza della traccia audio (non la dimensione del file) e verificare, anche in questo caso, che sia la stessa prima e dopo la conversione.

⁵¹ Cfr. AgID, *Linee guida sulla formazione, gestione e conservazione di documenti informatici*, cit., par. 1.11 (*Principi generali della gestione documentale*), p. 11.

⁵² A tal proposito l'*Allegato 2* precisa che «sebbene sia tecnicamente possibile continuare a produrre e riprodurre file in un formato deprecato, il progresso tecnologico, o per meglio dire l'obsolescenza tecnologica, condannerà tali file a non essere più definitivamente leggibili, a causa della mancanza di applicativi che ne implementino la decodifica (a meno di non coinvolgere risorse economiche solitamente sproporzionate al bisogno di accesso ai dati contenuti). Quando ciò accade ed è riconosciuto in maniera manifesta, si parla di formati "obsoleti"». Cfr. *Allegato 2*, par. 1.2.2 – *Classificazione dei formati*, p. 12.

più volte nella gestione del documento informatico e in diversi momenti per finalità gestionali o conservative»⁵³.

Lo strumento per valutare il livello di obsolescenza dei documenti è la cosiddetta “valutazione di interoperabilità”, che tutte le pubbliche amministrazioni devono effettuare almeno una volta l’anno⁵⁴. Sulla base del risultato di tale valutazione, potrebbe essere necessario pianificare il riversamento di documenti da un formato elettronico a un altro, al fine di garantirne la conservazione nel tempo⁵⁵. Inoltre, nella scelta di nuovi formati elettronici è importante tenere conto delle peculiarità tecniche del formato elettronico di origine e di destinazione, con particolare riferimento sia alla perdita di dati e metadati che alla diversa qualità o rappresentazione tecnica degli stessi; in alcuni casi, è obbligatoria la conservazione anche del documento nel suo formato originale.

L’Allegato 2 illustra in dettaglio i principi generali su cui deve basarsi il processo di riversamento di formato elettronico e le modalità tecnico-operative. Innanzitutto, il riversamento può avvenire operando su un singolo documento informatico alla

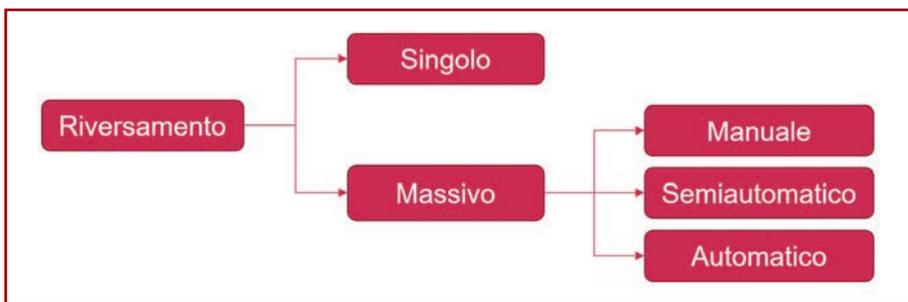


Figura 3. Le diverse tipologie di riversamento

⁵³ Ivi.

⁵⁴ La “valutazione di interoperabilità” è stata introdotta dalle citate *Linee guida*, secondo le quali «nello scegliere i formati di file [...] da utilizzare per i propri documenti informatici, i soggetti di cui all’art. 2 comma 2 e comma 3 del CAD [Codice dell’Amministrazione Digitale] possono effettuare una valutazione di interoperabilità che tenga conto dei seguenti fattori: formati aperti, non proprietari, standard de iure, estendibili, parlanti, completamente robusti, indipendenti dal dispositivo». La valutazione di interoperabilità «è effettuata in base alle indicazioni previste nell’Allegato 2, Formati di file e riversamento. La valutazione di interoperabilità, in quanto parte della gestione informatica dei documenti, viene effettuata periodicamente e, comunque, ogni anno, allo scopo di individuare tempestivamente cambiamenti delle condizioni espresse dai punti sopra elencati». Cfr. AgID, *Linee guida sulla formazione, gestione e conservazione di documenti informatici*, cit., par. 3.6 (*Formati di file*), p. 27-28. L’Allegato 2, al par. 3.1 (*Valutazione di interoperabilità*), elenca in dettaglio quali sono gli elementi che devono essere esaminati per condurre la “valutazione di interoperabilità”.

⁵⁵ «A seguito della valutazione di interoperabilità, i soggetti di cui all’art. 2 comma 2 e comma 3 del CAD [Codice dell’Amministrazione Digitale] valutano l’esigenza o l’opportunità di effettuare o pianificare il riversamento dei file da un formato di file ad un altro formato [...]. Il riversamento è effettuato in base alle indicazioni previste nell’Allegato 2 *Formati di file e riversamento*». Cfr. AgID, *Linee guida sulla formazione, gestione e conservazione di documenti informatici*, cit., par. 3.5 (*Riversamento*), p. 28.

volta (*riversamento singolo*) o su un insieme di documenti (*riversamento massivo*). In quest'ultimo caso, l'operazione di riversamento può essere compiuta in maniera *manuale, semiautomatica o automatica*⁵⁶ (Fig. 3).

Riguardo le motivazioni alla base del riversamento di formato, questo viene generalmente effettuato per finalità di conservazione (è il caso più frequente). Tuttavia, in taluni casi, può avere finalità di tipo gestionale⁵⁷, come nel caso della produzione di ulteriori versioni degli stessi documenti a scopo di esibizione o di fruizione da parte degli utenti (si pensi, ad esempio, al riversamento di un documento in formato DOCX nel formato PDF/A, effettuata per finalità di pubblicazione sul sito web di un Ente). Un aspetto da tenere in debita considerazione è quello della "destinazione" del documento originale una volta conclusa l'operazione di riversamento. Generalmente un riversamento di formato produce come effetto l'alterazione della sequenza di bit (*bitstream*) del documento, rendendo di fatto non più valide eventuali firme e sigilli apposti al documento. Di conseguenza, quando si effettua un riversamento finalizzato alla conservazione del documento «si può e, in certi casi previsti dalla legge, si deve, conservare anche la copia del file nel formato originario [...] purché sia conservata anche – in una forma logicamente e univocamente legata ad esso – copia conforme del medesimo in un formato adatto alla conservazione»⁵⁸. Ovviamente le stesse considerazioni si applicano non solamente ai formati di file, ma anche alle varie altre fattispecie (formati di buste, pacchetti di file, flussi binari, codec)⁵⁹. La *ratio* del mantenimento del documento nel formato originario è, quindi, riconducibile alla necessità di «consentire la convalida di firme, sigilli o validazioni temporali elettroniche che, eventualmente già presenti nel documento originario, non potrebbero essere riportate – senza invalidarle – nel documento riversato»⁶⁰. Appare evidente come, nel tempo, si possa presentare la necessità di conservare le varie e successive versioni di un documento sottoposto a vari riversamenti di formato, e questo potrebbe costituire una questione di non poca rilevanza nel lungo periodo⁶¹.

⁵⁶ La valutazione di interoperabilità comprende la redazione di un elenco dei «processi di riversamento di formato attualmente in corso nell'organizzazione, con particolare riferimento ai software applicativi impiegati e alle procedure tecniche (automatiche, semiautomatiche o completamente manuali) adottate per configurare tali riversamenti». Cfr. *Allegato 2*, par. 3.1 – *Valutazione di interoperabilità*, p. 147.

⁵⁷ Cfr. AgID, *Linee guida sulla formazione, gestione e conservazione di documenti informatici*, cit., par. 1.11 (*Principi generali della gestione documentale*): «Nell'ambito della gestione documentale possono essere necessarie attività di riversamento dei documenti in altro formato diverso da quello originale, come specificato al paragrafo 3.7. Tale riversamento può avvenire più volte nella gestione del documento informatico e in diversi momenti per finalità gestionali o conservative».

⁵⁸ Cfr. *Allegato 2*, par. 3.3 (*Riversamento*), punto 2.

⁵⁹ Per una definizione delle varie fattispecie si rimanda all'*Allegato 2*, par. 1.1.1 (*File, flussi digitali e buste-contenitori*), p. 4 e 5.

⁶⁰ *Ibidem*.

⁶¹ A quanti riversamenti di formato dovrà essere sottoposto un documento a conservazione illimitata nel tempo? E quante versioni dello stesso documento sarà necessario conservare?

Un altro aspetto molto importante è quello del profilo giuridico del documento riversato. A tal proposito l'Allegato 2 precisa che «quando si effettua un riversamento finalizzato alla conservazione, il file riversato è una copia digitale di un documento digitale⁶² e, come tale, la conformità della copia è attestata in base alla normativa vigente, inclusa la certificazione di processo»⁶³ di cui all'Allegato 3 delle *Linee guida AgID*.

Infine, è importante che «nella scelta dei nuovi formati di file (e, *mutatis mutandis*, di contenitori, pacchetti di file, flussi digitali e codec), così come nella scelta metodologica circa l'esecuzione del riversamento, si considerino le peculiarità tecniche del formato sorgente e di quello riversato, con particolare riferimento sia alla perdita di dati e metadati, sia alla diversa qualità o rappresentazione tecnica dei medesimi»⁶⁴.

Per quanto riguarda il *riversamento massivo* finalizzato alla conservazione di più documenti informatici – a prescindere se sia conservato o meno il documento nel suo formato originario – l'Allegato 2 precisa quali sono i principi da seguire:

- il riversamento di formato deve essere effettuato mediante un processo certificato in grado di garantire l'integrità e la riproducibilità del processo stesso; ogni procedura di riversamento deve essere descritta «in tutti i suoi dettagli tecnici, inclusi quelli sui formati di destinazione e quelli riversati»⁶⁵ nel manuale di gestione documentale ed eventualmente nella valutazione di interoperabilità;

- le pubbliche amministrazioni hanno l'obbligo di rispettare il principio generale⁶⁶ secondo il quale «il riversamento avviene sempre verso formati che ne migliorano l'interoperabilità, o comunque non la peggiorano (come stabilito, ad esempio, mediante il calcolo dell'indice di interoperabilità)»⁶⁷; in particolar modo non sono opportuni i riversamenti da un formato aperto verso formati chiusi⁶⁸; da un formato non proprietario, ovvero proprietario a libero utilizzo, verso formati proprietari; da un formato non dipendente dal dispositivo verso formati dipendenti dal dispositivo; da un formato parlante verso formati muti⁶⁹;

⁶² In realtà, si dovrebbe parlare di “copia informatica di documento informatico”, definita come «il documento informatico avente contenuto identico a quello del documento da cui è tratto su supporto informatico con diversa sequenza di valori binari» dall'art. 1, comma i-quater del *Codice dell'amministrazione digitale*.

⁶³ Cfr. *Allegato 2*, par. 3.3 (*Riversamento*), punto 3, p.150.

⁶⁴ Ivi, punto 4, p. 150.

⁶⁵ Ivi, punto 5, p. 150-154.

⁶⁶ Questo è un principio che si applica esplicitamente ai riversamenti di formato dei documenti informatici operati dalle Pubbliche Amministrazioni, ma essendo di tipo generale parrebbe opportuno estenderlo a tutti i soggetti, indipendentemente dalla categoria di appartenenza.

⁶⁷ Cfr. *Allegato 2*, par. 3.3 (*Riversamento*), p. 150.

⁶⁸ Si definiscono “aperti” quei formati le cui specifiche sono pubblicamente disponibili (eventualmente anche a pagamento, come avviene per molti formati che sono stati riconosciuti standard ISO). Al contrario, sono “chiusi” quei formati elettronici le cui specifiche non sono state rese pubbliche.

⁶⁹ Si definiscono “parlanti” quei formati elettronici che consentono l'inclusione di metadati al proprio interno (ad esempio, i formati JPEG e MP3: cfr. nota 45). Nel caso contrario, i formati si definiscono “muti”.

- per ogni documento informatico riversato con un processo massivo automatico, ovvero semiautomatico, il processo deve produrre un’*attestazione del riversamento* specifico di quel documento; le attestazioni relative ai documenti riversati come parte della medesima procedura devono essere collezionate in un “Registro di riversamento”;
- nel rispetto delle leggi sulla privacy in vigore, il riversamento in un altro formato costituisce un’ulteriore occasione per ottemperare «agli obblighi in materia di adeguatezza, pertinenza, minimizzazione ed esattezza dei dati personali ivi contenuti, così come della liceità del loro trattamento e della loro eventuale pseudonimia»⁷⁰;
- qualora sussistano obblighi di legge o altri tipi di vincoli che obblighino a preservare l’evidenza informatica (*bitstream*) del documento originale, quest’ultimo deve essere conservato insieme al suo riversamento in formato più interoperabile; in tal caso deve essere specificato come i due documenti (l’originale e il riversato) vengono logicamente associati fra loro (ad esempio, associando le due impronte crittografiche) e di tale associazione logica deve essere dato conto nel “Registro di riversamento”;
- di solito un riversamento di formato elettronico altera l’evidenza informatica, intaccandone dunque l’integrità da un punto di vista strettamente tecnico; tuttavia, esistono riversamenti di formato che mantengono sostanzialmente invariato il contenuto documentale; tali possibilità, qualora individuate, devono essere specificate nel manuale di gestione documentale, descrivendo «come il contenuto documentale viene sostanzialmente preservato, incluse le considerazioni sulla similitudine tra le due evidenze informatiche»⁷¹;
- in alcuni casi, l’imbustamento di un documento in un contenitore aggiuntivo mantiene integro il contenuto all’interno⁷²; quindi, un riversamento di questo genere può essere reversibile anche se l’integrità del documento nella sua interezza è compromessa; lo stesso succede quando si vuole inserire un documento all’interno di un pacchetto di file; entrambe le metodologie non risolvono i problemi di interoperabilità e obsolescenza originali per cui il riversamento è stato concepito, ma possono mitigarli, soprattutto se vengono scelti contenitori o pacchetti parlanti che compensano la mancanza di metadati nel formato originale;
- analogamente, quando un documento viene “reimbustato” (ovvero, il riversamento consiste solamente nell’inserire il documento in una nuova busta rispetto a quella che lo conteneva inizialmente), si potrebbe mantenere, in alcuni casi, l’integrità del contenuto ma non quella della busta in sé; questa metodologia potrebbe risolvere il problema di obsolescenza e interoperabilità legato al solo formato della busta usata in precedenza.

⁷⁰ Cfr. *Allegato 2*, par. 3.3 (*Riversamento*), p. 152.

⁷¹ *Ivi*, p. 153.

⁷² Ad esempio, è il caso di un documento inizialmente compresso in una busta ZIP che viene estratto e successivamente inserito in una busta RAR.

7. La certificazione di processo applicata all'operazione di riversamento di formato elettronico

La possibilità di ricorrere alla certificazione di processo⁷³ per le operazioni di conversione massiva di formato elettronico è stata esplicitamente prevista dall'Allegato 2 alle *Linee guida AgID*. Infatti, come già ricordato «quando si effettua un riversamento finalizzato alla conservazione, il file riversato è una copia digitale di un documento digitale e, come tale, la conformità della copia è attestata in base alla normativa vigente, inclusa la certificazione di processo come riportata nell'Allegato 3 delle presenti Linee guida»⁷⁴. Inoltre, quando si effettua un riversamento massivo finalizzato alla conservazione di più documenti informatici, il processo include una serie di passaggi che «possono essere inquadrati come parte di un processo certificato di conformità della copia riversata»⁷⁵.

Da quanto sopra si deduce che il documento informatico ottenuto a seguito di una conversione di formato elettronico costituisce una copia informatica di un documento informatico, ovvero un «documento informatico avente contenuto identico a quello del documento da cui è tratto su supporto informatico con diversa sequenza di valori binari»⁷⁶. Come tale, la conformità della copia deve essere attestata in base alla normativa vigente, in particolare secondo quanto previsto dall'art. 23-bis (*Duplicati e copie informatiche di documenti informatici*) del Codice dell'Amministrazione Digitale il quale stabilisce che le copie informatiche del documento informatico, se prodotte in conformità alle vigenti *Linee guida*, «hanno la stessa efficacia probatoria dell'originale da cui sono tratte se la loro conformità all'originale, in tutte le sue componenti, è attestata da un pubblico ufficiale a ciò autorizzato o se la conformità non è espressamente disconosciuta».

Si ricorda che nel caso della copia per immagine su supporto informatico di un documento analogico, le due metodologie possibili per assicurare che la copia abbia contenuto e forma identici a quelli del documento analogico da cui è tratto sono il raffronto dei documenti o la certificazione di processo nei casi in cui siano adottate tecniche in grado di garantire la corrispondenza della forma e del contenuto dell'originale e della copia⁷⁷. Estendendo questi concetti al caso dei documen-

⁷³ La certificazione di processo, prevista dal Codice dell'Amministrazione Digitale, art. 22 c. 1-bis e 23-ter c.1-bis, è dettagliata nell'Allegato 3 (*Certificazione di processo*) alle *Linee Guida sulla formazione, gestione e conservazione dei documenti informatici* dell'AgID. L'Allegato 3 fa riferimento al caso delle copie per immagine su supporto informatico di documenti analogici (Codice dell'Amministrazione Digitale, art. 1, comma 1, lettera i-ter). La certificazione di processo viene utilizzata tipicamente in contesti di digitalizzazione di grandi quantità di documenti cartacei, ma può essere adattata al caso della conversione di documenti informatici da un formato elettronico ad un altro.

⁷⁴ Cfr. Allegato 2, par. 3.3 (*Riversamento*), punto 3, p. 150.

⁷⁵ Cfr. Allegato 2, par. 3.3 (*Riversamento*), punto 5, p. 150-151.

⁷⁶ Cfr. *Codice dell'Amministrazione Digitale*, art.1, comma 1, lettera i-quater.

⁷⁷ Cfr. *Codice dell'Amministrazione Digitale*, art. 22, comma 1-bis.

ti digitali sottoposti ad un processo di riversamento di formato elettronico, è possibile ipotizzare che anche in questo contesto la conformità del documento riversato rispetto all'originale possa essere assicurata mediante la tecnica del raffronto (applicabile quando il numero di documenti informatici da convertire sia ridotto) oppure mediante la certificazione di processo (applicabile nel caso di riversamento massivo di documenti informatici).

Inoltre, mutuando la certificazione di processo applicata alla digitalizzazione di documenti analogici – come prevista dall'*Allegato 3* delle *Linee guida* – al caso della conversione di formato elettronico, è possibile ipotizzare che la conformità possa essere certificata secondo due modalità:

- l'attestazione da parte di un notaio o un pubblico ufficiale a ciò autorizzato; in tal caso il documento di attestazione, che prende il nome di "attestazione di conformità", viene sottoscritto con firma digitale o altra firma elettronica qualificata (come stabilito dagli art. 22 comma 2 e 23-ter comma 3 del Codice dell'Amministrazione Digitale), e ha un valore probatorio "privilegiato", ovvero fa piena prova fino a querela di falso (ex art. 2700 del Codice civile);
- l'attestazione da parte di un soggetto privato; in tal caso il documento di attestazione, che prende il nome di "rapporto di verifica", ha un valore probatorio "semplice" e fa piena prova solo fino a disconoscimento.

L'*Allegato 3* specifica che si parla propriamente di certificazione di processo solo nel primo caso, ovvero qualora l'attestazione di conformità venga rilasciata da notaio o pubblico ufficiale a ciò autorizzato e sottoscritta per mezzo della firma digitale o di altra firma elettronica qualificata.

Per quanto riguarda l'ambito oggettivo, il ciclo di conversione massiva dovrà essere certificato da organismo terzo in accordo agli standard ISO 9001 e ISO 27001, con campo di applicazione specifico per i servizi di progettazione e riversamento massivo di formato elettronico.

L'*Allegato 2* non fornisce indicazioni in merito alle fasi di cui potrebbe comporsi un ciclo di conversione; tuttavia, rimanda esplicitamente alla certificazione di processo descritta nell'*Allegato 3*, per cui è possibile immaginarlo come composto delle seguenti fasi:

- 1) descrizione accurata e completa dell'intero processo, da effettuarsi prima ancora che l'attività di riversamento massivo venga avviata sull'intero lotto di documenti da sottoporre a conversione; la procedura di riversamento «è descritta nel manuale di gestione e, se del caso [...], nella valutazione di interoperabilità, in tutti i suoi dettagli tecnici, inclusi quelli sui formati di destinazione e quelli riversati»⁷⁸;

⁷⁸ Cfr. *Allegato 2*, par. 3.3 (*Riversamento*), p. 151.

2) produzione della “certificazione iniziale” che attesti la conformità agli originali di un campione⁷⁹ di documenti sottoposti al processo di riversamento di formato elettronico, ricorrendo al tradizionale metodo del raffronto dell’originale con la copia. La certificazione iniziale definisce e “congela” i criteri di qualità e sicurezza del processo, così da assicurarli automaticamente nelle fasi successive. Da questo momento, questi criteri saranno validi per tutti i documenti che saranno sottoposti al processo di riversamento;

3) esecuzione del processo di riversamento, che viene effettuato utilizzando appositi strumenti software. Si faccia attenzione al fatto che dovrà essere effettuata la conversione non soltanto del formato elettronico del documento originale ma anche del set di metadati ad esso associato⁸⁰. In generale, gli strumenti di conversione tendono ad essere specifici per determinate tipologie di documenti (ad esempio, vi sono strumenti dedicati alla conversione delle immagini, altri che eseguono la conversione dei documenti sonori ecc.). Gli strumenti di conversione possono introdurre dei rischi, come le modifiche – più o meno “visibili” – che vengono apportate ai documenti durante il riversamento. Per ridurre questo rischio, sarebbe opportuno stabilire quali sono le “proprietà significative” che il documento riversato deve mantenere (ad esempio, in termini di formattazione, aspetto e funzionalità) e quelle che, invece, possono essere trascurate, e definire delle metriche per valutare la perdita di informazioni.

Come già anticipato nel paragrafo precedente, per ogni documento riversato di un processo massivo il processo di conversione deve produrre un’*attestazione del riversamento* specifico di quel documento; le attestazioni relative a documenti riversati come parte della medesima procedura devono essere collezionate in un “Registro di riversamento”, che contiene (globalmente rispetto alla particolare procedura e individualmente per ogni file elaborato) almeno i seguenti elementi⁸¹:

a) un riferimento temporale opponibile a terzi ai sensi dell’art. 41 del DPCM 22 febbraio 2013 relativo all’inizio o alla fine del riversamento (indicando chiaramente a quale dei due tempi ciascun riferimento temporale si riferisca);

b) indicazioni sul sistema informativo impiegato (per esempio: nome, numero di revisione sia del sistema operativo che del software; nome della macchina e suoi indirizzi di rete o altri numeri identificativi unici delle componenti hardware; identificativi unici del software quali i numeri di licenza; nome o identificativo unico e orario di accesso al sistema operativo dell’utenza sotto cui l’applicativo ha agito);

⁷⁹ Riguardo il piano di campionamento, l’Allegato 3 segnala la norma UNI ISO 2859 – *Procedimenti di campionamento nell’ispezione per attributi - Parte 1: Schemi di campionamento indicizzati secondo il limite di qualità accettabile (AQL) nelle ispezioni lotto per lotto*.

⁸⁰ Su questo punto, si vedano le considerazioni sui metadati, specialmente quelli interni, presenti nell’Allegato 2, par. 3.3 (*Riversamento*), p. 153.

⁸¹ Cfr. Allegato 2, par. 3.3 (*Riversamento*), p. 151-152.

- c) il nome del documento sorgente, la sua posizione nel filesystem e i metadati esterni;
 - d) il formato sorgente del documento, la versione del formato⁸² e i metadati interni;
 - e) l'impronta crittografica del documento sorgente;
 - f) il nome del documento riversato, la sua posizione nel filesystem e i metadati esterni;
 - g) il formato di riversamento, la sua versione e i metadati interni convertiti dal punto d;
 - h) l'impronta crittografica del documento destinazione;
 - i) nel caso di formati "contenitori" (come sorgente o destinazione), i metadati ai punti d, g si intendono riferiti alla busta, cui si aggiunge elenco completo del contenuto della medesima (per esempio, le essenze e i codec impiegati per ciascuna di esse, con i loro eventuali metadati e profili);
 - j) nel caso di formati di "pacchetto" (come sorgente o destinazione), i controlli c-h si intendono riferiti a ciascun oggetto digitale componente il pacchetto, cui si aggiunge l'indicazione dei metadati deducibili dall'intero pacchetto nella sua interezza, laddove non esplicitamente descritti negli eventuali file-manifesto;
 - k) eventuali errori tecnici, anomalie o ambiguità riscontrate durante il riversamento;
- 4) produzione della "certificazione di chiusura": il ciclo di riversamento dovrà concludersi con il metodo del raffronto su una *campione* di documenti riversati, generando una "attestazione di conformità" nel caso sia sottoscritto da un notaio o un pubblico ufficiale oppure un "rapporto di verifica" nel caso vi sia l'intervento di un soggetto privato.

8. La tutela archivistica

È bene ricordare che gli interventi di riversamento di formato elettronico sono sottoposti al regime di tutela da parte delle competenti strutture del Ministero della Cultura. Infatti, le disposizioni del *Codice dei beni culturali e del paesaggio* prevedono l'obbligo per gli archivi delle pubbliche amministrazioni e per quelli privati dichiarati di interesse storico particolarmente importante ai sensi dell'art. 13, di ottenere l'autorizzazione sia per quanto riguarda l'esecuzione degli interventi stessi, sia in merito alla (eventuale) eliminazione degli originali di cui si sia effettuato il riversamento. In particolare, ai sensi dell'art. 21, commi 4 e 5⁸³, le

⁸² Ad esempio, nel caso di un documento in formato PDF, occorre specificare quale sia la versione (le versioni che si sono succedute fino ad oggi sono le seguenti: PDF 1.0, PDF 1.1, PDF 1.2, PDF 1.3, PDF 1.4, PDF 1.5, PDF 1.6, PDF 1.7, PDF 2.0). Considerazioni analoghe valgono per gli altri formati elettronici.

⁸³ Cfr. D.lgs. 22 gen. 2004, n. 42, *Codice dei Beni culturali*. L'art. 21 (*Interventi soggetti ad autorizzazione*), comma 4 stabilisce che «l'esecuzione di opere e lavori di qualunque genere su beni culturali è subordinata ad autorizzazione del soprintendente». Il successivo art. 5 precisa che «l'autorizzazione è resa su progetto o, qualora sufficiente, su descrizione tecnica dell'intervento, presentati dal richiedente, e può contenere prescrizioni. Se i lavori non iniziano entro cinque anni dal rilascio dell'autorizzazione, il soprintendente può dettare prescrizioni ovvero integrare o variare quelle già date in relazione al mutare delle tecniche di conservazione».

operazioni di riversamento di formato elettronico che hanno come oggetto i documenti di archivi pubblici o di quelli privati riconosciuti di interesse storico particolarmente importante, rientrano tra le attività soggette ad autorizzazione da parte delle strutture competenti del Ministero della Cultura (ad esempio, le Soprintendenze archivistiche e bibliografiche competenti per territorio nel caso di archivi di enti pubblici non statali o di archivi privati dichiarati). Allo stesso modo l'eventuale eliminazione dei documenti originali dopo che sono stati sottoposti al processo di riversamento di formato elettronico è a tutti gli effetti un'operazione di "scarto" e come tale rientra tra le attività sottoposte ad autorizzazione, ai sensi del Codice dei Beni culturali, art. 21, comma 1, lettera d)⁸⁴.

Allo stato attuale, non sembra esserci una sufficiente consapevolezza su questi obblighi né da parte dei soggetti detentori di archivi, né da parte dell'Amministrazione archivistica dalla quale ci si aspetta un inevitabile intervento in questo ambito.

9. Considerazioni conclusive

Pur essendo stato teorizzato come una delle strategie di conservazione digitale più efficaci, occorre riconoscere che il riversamento di formato non viene ancora sufficientemente praticato e anche quando si avverte la necessità di convertire documenti digitali da un formato obsoleto a uno più moderno, l'operazione viene spesso ritardata a causa della mancanza di conoscenze e competenze, nonché di linee guida e raccomandazioni tecnico-operative su come effettuare la conversione e su quali strumenti utilizzare. Purtroppo, anche le *Linee guida sulla formazione, gestione e conservazione dei documenti informatici*, più volte citate, non hanno risolto il problema, perché non forniscono indicazioni concrete sul processo di conversione (ad esempio: quali strumenti software utilizzare, quali metriche adoperare per misurare la perdita di informazioni e testare l'accuratezza e la qualità delle conversioni ecc.). Inoltre, preoccupa la mancanza di strumenti software di conversione "certificati" dall'AgID che diano garanzie sui risultati ottenibili e ai quali affidare i processi di riversamento. Inoltre, vi sono delle questioni che destano preoccupazione e devono essere ulteriormente approfondite, come la perdita di contenuto "informativo" che potrebbe verificarsi a seguito di riversamenti ripetuti nel tempo⁸⁵, oppure la previsione della

⁸⁴ Ivi. L'art. 21, comma 1, lettera d) stabilisce che è subordinato ad autorizzazione del Ministero «lo scarto dei documenti degli archivi pubblici e degli archivi privati per i quali sia intervenuta la dichiarazione ai sensi dell'articolo 13, nonché lo scarto di materiale bibliografico delle biblioteche pubbliche [...] e delle biblioteche private per le quali sia intervenuta la dichiarazione ai sensi dell'articolo 13».

⁸⁵ È per questo che all'atto della formazione del documento informatico occorre preferire quei formati elettronici che diano maggiori garanzie di longevità, al fine di minimizzare il più possibile il numero di riversamenti che si renderanno necessari nel tempo.

necessità di conservare tutte le “copie informatiche” nei vari e successivi formati che verranno prodotte nel tempo⁸⁶.

Stante questa situazione, è facile comprendere il motivo per cui, al momento attuale, non si è a conoscenza di sistemi di conservazione digitale che abbiano già implementato le funzionalità necessarie per avviare operazioni di conversione, almeno nel contesto italiano. Tuttavia, in molti depositi digitali vi sono già numerosi documenti codificati in formati ormai obsoleti e a rischio di non poter più essere letti – come quelli acquisiti alla fine degli anni ‘90 del secolo scorso⁸⁷ – e che, quindi, devono essere riversati con una certa urgenza. Non è più possibile attendere oltre ed è giunto il momento di cominciare ad avviare le prime operazioni di riversamento. Per governare questi processi servono figure professionali altamente specializzate e formate, preferibilmente mediante percorsi post-universitari incentrati sulle tematiche degli archivi digitali che possano fornire le conoscenze e le competenze necessarie. Inoltre, sarebbe altamente desiderabile che la comunità internazionale degli archivisti, in particolare coloro che sono impegnati sui temi della *digital preservation*, si dedicasse con maggiore attenzione a queste problematiche e intraprendesse uno sforzo collaborativo finalizzato allo sviluppo di linee guida e raccomandazioni condivise.

The problem of file format obsolescence poses significant challenges for the long-term preservation of electronic records. Continuous technological developments make existing formats obsolete, requiring timely conversion to new formats to ensure future usability. This issue affects both open-source and commercial formats, with vendors sometimes encouraging a switch to new products through planned obsolescence. To solve this problem, the strategy of file format conversion has been proposed, which involves converting electronic records from an old format to a more current one. However, although the importance of this strategy is widely recognised, its practical implementation is still limited due to a lack of knowledge and expertise, operational guidelines, and advice. This paper aims to fill this gap by outlining the rationale behind format conversion, the timing of its implementation, and the methodologies for conversion – referencing both international and national guidelines – and proposing a methodology for mass conversion based on process certification.

⁸⁶ Cfr. la nota 62.

⁸⁷ Per fare un esempio, si pensi ai documenti di testo creati con le vecchie versioni di Microsoft Word (a partire da Word per Windows 1.0 fino a Word per Windows 2003, che utilizzavano il formato DOC nelle sue varie versioni) e che necessitano di essere convertiti almeno nel formato DOCX.

L’ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

La trascrizione automatica di documenti a stampa antichi. Appunti per un modello di riconoscimento della tipografia in corsivo

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00094

Stefano Bazzaco

Università di Verona

Il contributo intende descrivere il flusso di lavoro che ha condotto alla produzione di un modello di Handwritten Text Recognition (HTR) per la trascrizione automatica di testi veneziani del Cinquecento in corsivo. Nella prima parte, si definisce l'ambito di studio, con particolare attenzione allo stato dell'arte e ai recenti sviluppi nel campo dell'HTR per quanto riguarda i complex scripts, vale a dire testi a stampa antichi e manoscritti che per le loro caratteristiche ostacolano l'applicazione dei tradizionali sistemi di Optical Character Recognition (OCR). Nella seconda parte, si espongono le principali fasi del lavoro di addestramento della macchina per la creazione del modello Italics_VeniceXVIs, che costituisce un primo passo per l'interpretazione dei testi cavallereschi in carattere corsivo di interesse del Progetto Mambrino dell'Università di Verona. Infine, si individuano le principali caratteristiche del modello e, in un'ottica di accessibilità e riutilizzo, si segnalano i passi futuri del progetto, suggerendo possibili ricadute della ricerca svolta in relazione con altri ambiti di studio.

Introduzione

Negli ultimi anni il processo di digitalizzazione del patrimonio bibliografico sta sperimentando degli avanzamenti notevoli, che a loro volta si ripercuotono sui metodi e le pratiche degli addetti ai lavori, vale a dire umanisti, bibliotecari e in generale esperti della gestione dell'informazione. Tra i vari strumenti digitali che si impiegano per la conversione, rimediazione e analisi di materiali analogici testuali, stanno destando particolare attenzione le applicazioni che sfruttano l'intelligenza artificiale per svolgere compiti che richiederebbero altrimenti il dispendio di un'enorme quantità di tempo e di forza lavoro da parte degli specialisti. Si tratta di sistemi di *machine learning* che per risolvere particolari compiti integrano alla potenza di calcolo della macchina l'apprendimento su base di dati fornita dall'utente, aprendo la via a possibilità predittive che promettono di avere profonde ricadute sul trattamento stesso dei dati informativi e, conseguentemente, della conoscenza. Basti considerare, per esempio, la recente elaborazione di tecnologie come i

Large Language Models (LLM), modelli linguistici che, appoggiandosi a reti neurali per svolgere funzioni generative anche molto complesse, stanno scuotendo le fondamenta della ricerca e della didattica perché impongono di ripensare il patrimonio culturale come una sorta di *big data*, suscettibile di analisi quantitative per l'estrapolazione, la produzione e la valorizzazione automatica dell'informazione¹.

Precedentemente allo sviluppo dei LLM, l'ambito di ricerca maggiormente esposto alle nuove proposte dell'intelligenza artificiale è stato quello della trascrizione automatica, un campo di studi da troppo tempo imbrigliato dalle limitazioni e i *bias* dell'insufficienza tecnologica. Nello specifico, l'introduzione di reti neurali ricorrenti, vale a dire processi che simulano il funzionamento della mente umana ai fini di addestrare la macchina a svolgere funzioni complesse², ha avuto un notevole impatto sulle applicazioni di Automatic Text Recognition (ATR), modificandone le prestazioni e suggerendone un maggior impiego su larga scala. Grazie alla preparazione e distribuzione di modelli di riconoscimento di testo altamente affidabili, infatti, si assiste a una proliferazione di progetti che mirano a rendere accessibile e processabile il patrimonio testuale globale avvalendosi di strumenti di trascrizione automatica che consentono di trasformare il testo contenuto nelle immagini in formato elettronico (*machine readable*).

Tra le tendenze più significative che si registrano nel contesto della ATR, si segnala il vertiginoso sviluppo di sistemi di Handwritten Text Recognition (HTR), che hanno rapidamente soppiantato in termini di impiego i più famosi, ma datati, sistemi di Optical Character Recognition (OCR). Questa sostituzione ha in qualche modo alimentato l'erronea opinione che l'HTR sia l'evoluzione naturale dell'OCR, quando invece si tratta di tecnologie in parte distinte, visto che la più antica, l'OCR, si basa sul riconoscimento delle realizzazioni grafiche di singoli caratteri, mentre la più recente, l'HTR, cerca di individuare *pattern* ricorrenti di segni grafici in una linea di testo, affidandosi alla predizione dei risultati di trascrizione grazie alle informazioni contestuali e all'addestramento preliminare fornito dall'utente.

L'allenamento della macchina, che si avvale di processi di *machine learning* come le reti neurali ricorrenti (RNN), costituisce la principale caratteristica dei sistemi di HTR ed ha favorito un significativo miglioramento dei risultati di trascrizione automatica, sia nel caso di testi manoscritti che di testi a stampa. Di conseguenza, la riflessione circa l'impiego di questi strumenti per la digitalizzazione del patrimonio culturale testuale si sta spostando verso problematiche specifiche, come la riduzione del lavoro umano nella creazione del materiale di addestramento (*training set*) e il riuso di dati e modelli già esistenti.

¹ Nell'attualità esistono diversi modelli linguistici generativi: il più comune e diffuso, per ragioni di distribuzione e risonanza mediatica, è ChatGPT sviluppato da Open AI, ma altri LLM sono stati resi pubblici negli ultimi mesi, come Bard (Google), Claude 2 (Anthropic) e LLAMA (Meta). Per uno studio approfondito delle caratteristiche e delle possibilità che potrebbero offrire queste applicazioni alla ricerca umanistica si vedano Ciotti 2023; Roncaglia 2023.

² Schwartz-Ricci 2022, p. 172.

Il presente studio si colloca in questo contesto e mira a descrivere le principali scelte che hanno governato il progetto di creazione di un modello di HTR per l'interpretazione dei caratteri in corsivo del Rinascimento. Inizialmente, si propone un'indagine storica dei sistemi di HTR, ponendo l'accento sullo stato dell'arte e il flusso di lavoro. In secondo luogo, si presentano le piattaforme disponibili e si definiscono le fasi progettuali che caratterizzano la creazione di un modello di riconoscimento, con riferimento in particolare al riutilizzo dei dati, allo sviluppo di modelli HTR misti e all'operazione di *fine tuning* di modelli preesistenti. Segue quindi una descrizione delle principali attività nel campo della trascrizione automatica intraprese dal gruppo di ricerca Progetto Mambrino dell'Università di Verona nell'ambito del Progetto PRIN "Mapping Chivalry" (2018-2022), del Progetto di Eccellenza "Inclusive Humanities" (2023-2027) e del Centro dipartimentale DAIH (Digital Arena for Inclusive Humanities). Infine, si chiariscono le caratteristiche dei modelli di HTR creati per l'interpretazione del corsivo, suggerendo possibili prospettive di ricerca future che possano favorirne il perfezionamento e la distribuzione su larga scala.

Trascrizione automatica e sistemi di HTR: breve analisi storica

La trascrizione automatica è un campo di studi che sta destando interesse grazie all'impressionante incremento delle prestazioni di calcolo che offrono le nuove tecnologie. In particolare, i sistemi di HTR, impiegati in larga scala per la conversione di intere collezioni digitalizzate in testo elettronico, hanno sperimentato un notevole sviluppo negli ultimi anni, qualificandosi come uno dei settori in maggior espansione nell'ambito delle Digital Humanities.

In origine, con il termine HTR si fa riferimento a strumenti che a partire dall'interpretazione di un'immagine digitalizzata di un manoscritto consentono di ottenere come *output* la trascrizione del contenuto del documento in formato testo elettronico. Per questa ragione, spesso ci si riferisce a queste tecnologie con la denominazione di "OCR for manuscripts", utilizzando un'espressione che risente di un certo grado di inesattezza perché accomuna due sistemi di riconoscimento distinti e al contempo limita l'applicabilità di entrambi a contesti predeterminati. E se, da una parte, è evidente che le rispettive storie dei sistemi di OCR e HTR si intreccino e sovrappongano, con il risultato che spesso non ci siano differenze sostanziali rispetto ai metodi e procedimenti alla base del loro utilizzo (Stokes e Kiessling 2024), dall'altro vanno chiarite le specificità dei secondi per spiegare l'interesse che stanno suscitando e il loro impiego sempre più diffuso.

La prima differenza è legata alle modalità di interpretazione dei segni grafici presenti nella fonte. Al riguardo, si noterà come già a livello terminologico gli strumenti di OCR si basano sulla decodificazione di caratteri isolati (Optical *Character* Recognition), mentre quelli di HTR su porzioni di testo (Handwritten *Text* Recognition), vale a dire linee o *patterns* di segni grafici contigui. Come naturale

conseguenza, si tende a considerare che i primi siano adatti all'interpretazione di testi a stampa, vista la loro predisposizione a trattare fenomeni unitari e discreti; al contrario, i secondi risponderebbero all'esigenza di trascrivere testimoni manoscritti, in cui lettere e parole tendono a essere collegate tra loro in modo asistematico, dipendendo da vari fattori, come le specificità della grafia dell'autore o le circostanze di scrittura. Questa distinzione però risulta immotivata, perché i sistemi di HTR vengono impiegati anche per la trascrizione automatica di testi a stampa antichi (ad esempio un libro in caratteri gotici) o di difficile interpretazione (ad esempio i giornali tedeschi dei primi del Novecento in Fraktur), raggiungendo un grado di affidabilità impensabile per i sistemi di OCR coetanei.

In secondo luogo, seppur gli sviluppi nel campo dell'OCR interessino diverse tecnologie gemelle, la storia dell'HTR si articola in modo parzialmente indipendente. Tra il 1965 e il 1973, quando si stavano già diffondendo i sistemi hardware di OCR di seconda generazione³, la trascrizione automatica dei testi manoscritti iniziava a muovere i primi passi e occupare le riflessioni di Sayre. Lo studioso segnalava le difficoltà legate all'interpretazione di scritture continue e proponeva come soluzione tecnica di regolarizzare i caratteri che compongono una parola, con l'intento di semplificare la segmentazione e il posteriore riconoscimento con tecniche di *template matching*⁴. Queste teorie trovarono scarsa applicazione, e bisognerà attendere sino agli anni '80 perché i sistemi di riconoscimento di testo vengano per la prima volta impiegati per identificare brevi scritture manoscritte, come codici postali e sequenze di assegni bancari, vale a dire materiali specifici dove la segmentazione era normalmente agevolata dalla disposizione ben separata dei caratteri.

Il quadro rimaneva poi essenzialmente immutato per un decennio, con la ricerca che si orientava verso lo sviluppo di dispositivi hardware per il riconoscimento gestuale e della grafia per sostituire tastiera e mouse nell'interazione uomo-macchina: questo campo, comunemente riconosciuto come *pen computing*, in quanto implica l'uso di una penna digitale e di una tavoletta grafica per l'elaborazione del segnale analogico in ingresso, si sviluppò a partire dal 1983, quando l'azienda Pencept elaborò prototipi come il PenPad 200 e il PenPad 320, con il secondo già integrato alle applicazioni MS DOS dei primi personal computer di IBM.

In seguito, tra gli anni '80 e '90, si diffusero i primi sistemi software per il riconoscimento di testo manoscritto, ma il limite principale di questi strumenti continuava ad essere quello di seguire un approccio *segmentation based*, dove l'interpreta-

³ Bazzaco et al. 2022, p. 73-74.

⁴ La questione di cui si interessa lo studioso prende il nome di "paradosso di Sayre" e si appoggia sulla seguente concezione: una parola manoscritta per essere interpretata deve essere suddivisa nei caratteri che la compongono, ma, inevitabilmente, per essere segmentata, la stessa dovrà essere prima interpretata. Come risoluzione del problema, lo studioso propone di suddividere in frammenti una linea di testo continua (segmentazione implicita) tentando di individuare una corrispondenza su base probabilistica tra un set di parole regolarizzato e gli stessi frammenti. Sull'argomento si veda: Sayre 1973 e Vinciarelli 2003.

zione passava per l'isolamento dei caratteri e la loro successiva interpretazione per sovrapposizione con un set predefinito di segni grafici. Questa disposizione presupponeva il ricorso a modelli euristici e algoritmi *ad hoc*, con evidenti ricadute sulla scalabilità, perché implicava la dipendenza del software dai dati di origine e una minore robustezza dell'applicazione con nuovi materiali da processare⁵.

Solo nel 2009, con l'introduzione di processi di *machine learning* che simulano l'attività umana, si assiste al superamento del paradigma iniziale e all'evoluzione progressiva e generalizzata dei sistemi di ATR. Gli studiosi, appoggiandosi sulle nuove tecnologie, intendono sfruttare la funzione di memoria predittiva che sta alla base delle reti neurali ricorrenti (RNN), come i processi Long Short-Term Memory (LSTM), per accrescere l'affidabilità dei software di trascrizione automatica. Gli strumenti che traggono maggiore vantaggio dallo sviluppo di questi componenti sono proprio i sistemi di HTR, poiché le loro prestazioni si basano su un addestramento manuale che potenzialmente deve consentire di trascrivere testi sconosciuti, che non sono stati precedentemente processati dalla macchina.

Le caratteristiche proprie dei sistemi di *machine learning* incidono anche sulle modalità di allenamento del software di HTR, in quanto il set di dati fornito dall'utente rappresenta il nucleo dell'intero processo e va confezionato seguendo modalità che ne assicurino successivamente l'efficienza e l'esportabilità a contesti diversi. Secondo questa prospettiva, il lavoro di trascrizione manuale deve rispettare dei criteri che assicurino un certo grado di robustezza del modello (*consistency*), il quale occasionalmente può essere sottoposto a operazioni di messa a punto (*fine tuning*). Così come avviene per i LLM, che su questo aspetto somigliano molto ai sistemi di HTR, l'addestramento della macchina è tanto più efficiente quanto più si incrementa la quantità e qualità dei materiali processati; ciò nonostante, la varietà interna a volte si deve preferire alla costituzione di un dataset molto esteso di materiali dello stesso tipo, ai fini di limitare l'iperspecializzazione del modello.

Ciò che accomuna le recenti tecnologie di Intelligenza Artificiale è proprio la necessità di un addestramento su materiali di base forniti dall'utente (*ground truth*) che rappresentano il modello ideale. Nel caso di strumenti come ChatGPT, la creazione di *ground truth* è un'operazione costosa e faticosa che richiede la revisione di miliardi di parole. Nel caso della trascrizione automatica, invece, i dati immessi dall'utente possono essere di quantità nettamente inferiore, ma non per questo si riducono i costi e le implicazioni etiche dell'operazione. In entrambi i casi descritti, infatti, il materiale di allenamento, che può essere considerato una sorta di big data testuale, è frutto di una selezione operata dall'utente, che così alimenta lo sviluppo della nuova tecnologia. Ma la scelta dei materiali processabili chiaramente è motivata da trend che governano la cultura dominante, per cui si preferisce lavorare su oggetti culturali di enorme interesse, più accattivanti per la disseminazio-

⁵ Vinciarelli 2003, p. 8-9.

ne, la collaborazione e la ricerca⁶. Questo comporta il rischio che interi patrimoni culturali già scarsamente visibili diventino totalmente opachi, mentre altri, che già stavano al centro dell'attenzione ed erano disponibili in grande quantità, accrescano a dismisura la loro visibilità e disponibilità, rinsaldando così la posizione culturale dominante di chi gestisce quegli stessi materiali. Secondo questa prospettiva, la creazione di *ground truth* per alimentare sistemi di *machine learning* è un'operazione nient'affatto banale, che deve necessariamente confrontarsi con i concetti di accessibilità ed equità nella produzione e riuso digitale dei beni culturali, in linea con le disposizioni della nuova agenda europea in materia di digitalizzazione.

Sistemi di HTR: stato dell'arte e flusso di lavoro

I sistemi di HTR, per le ragioni fin qui esposte, stanno suscitando l'interesse degli studiosi di materie umanistiche perché assicurano un grado di affidabilità elevato nella trascrizione automatica di documenti antichi. Seguendo questa prospettiva, il loro impiego è vantaggioso soprattutto con documenti a stampa antichi, perché spesso queste tecnologie consentono di ottenere un testo elettronico che si avvicina al 99% di precisione (vale a dire 99 caratteri corretti su 100, spazi inclusi). Si tratta di risultati impensabili fino a qualche decennio fa, quando gli strumenti di OCR con i medesimi documenti promettevano risultati meno attendibili (con un margine di precisione che si aggirava intorno all'80-85% nel migliore dei casi), alimentando una certa reticenza o avversione da parte degli specialisti del testo (*bias*), i quali ne segnalavano la completa inutilità per studi di carattere accademico.

Nell'attualità, invece, il panorama è sostanzialmente mutato, perché i sistemi di HTR permettono di risolvere, generalmente senza ostacoli, problemi prima insormontabili, come la trascrizione automatica in presenza di legature, abbreviature e caratteri speciali. Questi segni grafici, infatti, seppur facilmente interpretabili da parte di un lettore umano, condizionavano i risultati del riconoscimento con software di OCR, abbassandone drammaticamente le prestazioni. Lo stesso discorso si potrebbe estendere anche ai documenti manoscritti, che tuttavia costituiscono una sfida ancora maggiore per le nuove tecnologie perché presentano un'elevata variabilità interna anche a livello di realizzazione grafica di singole lettere.

Sulla scorta delle inedite potenzialità offerte dai sistemi di HTR attuali, sono stati inizialmente i grandi gruppi editoriali e le aziende tecnologiche private a integrare questi strumenti tra i loro servizi. Adam Matthew Digital, per esempio, impiega gli HTR per offrire all'utente una base di ricerca all'interno delle proprie collezioni digitalizzate; Google, invece, ha sviluppato di recente la piattaforma *Fabricius*, che si serve dell'IA per la trascrizione automatica di geroglifici. In questi casi, però, non si forniscono dichiarazioni riguardo gli algoritmi impiegati e non si diffonde il codice sorgente, pertanto l'implementazione resta appannaggio di una comunità ridotta di utenti⁷.

⁶ Stokes e Kiessling 2024.

⁷ Terras 2022, p. 184.

Sul versante opposto si situano invece le piattaforme di HTR sviluppate nel contesto di una comunità scientifica più estesa e che si caratterizzano per essere dei *Virtual Research Environment*, vale a dire delle applicazioni collaborative che possono essere utilizzate da qualsiasi tipologia di utente per finalità specifiche.

Questi strumenti propongono un flusso di lavoro che si compone di varie fasi:

1. *Data collection e preprocessing delle immagini*. In primo luogo, si procede con la definizione dei materiali che costituiranno la base dell'allenamento del modello HTR. Questi possono essere estratti da un unico documento (*single document*) o da documenti diversi (*multiple documents*), accomunati tra loro dal fatto di presentare caratteristiche linguistiche, grafiche o tipologiche condivise. In occasione, questi stessi materiali, prima di essere importati nella piattaforma, possono essere processati in locale con software che consentono di migliorarne la leggibilità, riducendo il rumore causato da elementi, come macchie o trasparenze di inchiostro, che complicano il riconoscimento. Nel caso dei sistemi di OCR, la critica sottolinea come questa operazione sia indispensabile ai fini di ottenere risultati di segmentazione e trascrizione potenzialmente affidabili: nella maggior parte dei casi per limitare il problema del rumore presente nelle digitalizzazioni si sceglie di convertire l'immagine in bianco e nero (*byinarization*). I sistemi di HTR recenti, al contrario, integrano applicazioni che preferiscono immagini a colori ad altissima risoluzione, in cui la macchina è capace di determinare le zone di testo presenti nella pagina a partire dalla distribuzione dei pixel, individuando e separando tra di loro aree di maggiore e minore densità di colore.

2. *Segmentazione*. Una volta importate le immagini digitalizzate all'interno della piattaforma, si compie una prima operazione fondamentale nel processo di trascrizione automatica che prende il nome di *Layout Analysis* e corrisponde alla segmentazione della pagina in regioni, righe di testo e, se necessario, parole. Normalmente, un tool di HTR integra tra le sue funzionalità un sistema di segmentazione automatica abbastanza affidabile, che distingue principalmente la/e cassa/e di testo e le corrispondenti linee che la/e compongono, mentre si trascura solitamente la suddivisione di linee in entità minori come le parole, poiché, come già chiarito, il riconoscimento si basa su *pattern* di segni grafici ricorrenti distribuiti in una linea di testo. Successivamente, l'utente ha la possibilità di correggere o ritoccare la segmentazione manualmente per mezzo di appositi strumenti integrati alla piattaforma.

3. *Creazione della ground truth*. Si tratta di una delle fasi fondamentali del flusso di lavoro e corrisponde alla corretta segmentazione e trascrizione manuale (secondo criteri prestabiliti) di alcune pagine di testo, che costituiscono il materiale di base (*golden standard*) su cui si fonda l'allenamento del software. L'importanza di questa attività è determinata da due ragioni principali: da un lato, deve rispettare criteri di qualità e coerenza metodologica poiché è la base per la creazione di un modello di HTR solido (*consistent*); dall'altro, rappresenta il nucleo del lavoro di

ricerca, e pertanto deve essere sottoposta ad una valutazione per la preservazione a lungo termine dei dati, secondo formati standard che ne consentano la futura esportazione ad altri contesti. Inoltre, questa fase può avvalersi di strumenti che facilitano la trascrizione manuale, per esempio si potrebbe prevedere l'impiego di modelli HTR già disponibili ai fini di generare una trascrizione preliminare dei contenuti, che, sebbene imprecisa, serva da base al ricercatore per la creazione del dataset finale⁸.

4. *Allenamento e creazione di un modello HTR.* Una volta realizzata la trascrizione di un numero adeguato di pagine (la cui quantità dipende dalle caratteristiche del testo sottoposto a riconoscimento), si procede con l'avvio della fase di *training*, vale a dire l'addestramento vero e proprio della macchina. Di solito l'allenamento prevede la suddivisione della *ground truth* in due parti: la prima, corrispondente all'insieme di allenamento (*training set*), è composta dal 90% del dataset e costituisce la base concreta dell'addestramento; la seconda, conforme all'insieme di validazione (*validation set*), è composta dal 5-10% del materiale trascritto dall'utente e permette di valutare l'efficienza del modello ottenuto. L'addestramento si articola in fasi che prendono il nome di *epochs*, vale a dire periodi successivi in cui l'applicazione tenta ricorsivamente di "avvicinarsi" alla trascrizione modello del *validation set* basandosi sulle informazioni acquisite durante il *training*. A tal riguardo, è bene considerare che non sempre la scelta di un maggior numero di *epochs* corrisponde a una miglior produttività: un modello di HTR, infatti, può tendere all'iperspecializzazione sui materiali di training, perdendo in efficienza se applicato a materiali affini. Questo fenomeno di atrofizzazione del modello si conosce come *overfitting* e viene normalmente controbilanciato attraverso la funzione di *early stopping*, che implica l'interruzione dell'addestramento se non si verifica un progressivo avvicinamento tra gli indici di errore di *training* e *validation set*⁹.

5. *Valutazione del modello.* L'indice di errore del modello, in inglese Character Error Rate (CER), viene calcolato determinando la distanza tra la trascrizione corretta fornita dall'utente e la trascrizione generata dal software. In concreto, questo numero percentuale viene stabilito in base alla sommatoria delle operazioni (aggiunte, eliminazioni, sostituzioni) che consentono di ottenere il testo corretto a

⁸ Al riguardo, ricerche attuali segnalano possibili soluzioni per produrre una *ground truth* sintetica, per esempio Perdiki 2023.

⁹ Su questi temi si veda Rabus, che afferma: «Using many epochs typically leads to the asymptotic approximation of the training data CER curve to zero, which means that the model has more or less completely adapted to the training data. However, while the fact that using a great number of epochs during training takes up many computational resources is getting less and less significant nowadays due to improved hardware and software capabilities, another important unintended effect brought about by using many training epochs must not be neglected. Since more training epochs adapt the model to the training data better and better, it may happen that model performance on test data stagnates or even becomes worse. This phenomenon is called overfitting and is quite common when training neural networks» (2019, p. 12).

partire dal testo grezzo trascritto dalla macchina¹⁰. Da ciò deriva che l'indice non corrisponde a un dato assoluto, ma relativo, perché fondato su un dataset di prova che non include testi totalmente estranei al modello, ma documenti spesso impiegati per lo stesso addestramento. Per questa ragione, il CER è un indicatore a volte fuorviante, che va rivalutato ogni volta che si sottopongono al modello nuovi materiali da interpretare.

Sviluppi recenti: piattaforme disponibili, modelli di HTR misti e fine tuning

Per svolgere le operazioni descritte all'interno di un unico spazio di lavoro, il ricercatore ha a disposizione due piattaforme ad accesso aperto: Transkribus (READ Coop)¹¹ ed eScriptorium¹². Si tratta delle due applicazioni più famose e utilizzate nell'ambito della ricerca universitaria e archivistica, perché entrambe dispongono di un'interfaccia grafica intuitiva e non richiedono competenze nell'ambito della programmazione. Inoltre, entrambi i software seguono grossomodo il flusso di lavoro descritto nei precedenti paragrafi, consentendo la creazione di modelli di HTR a partire dai dati di specifico interesse dell'utente.

La differenza principale tra Transkribus ed eScriptorium si basa sulle prospettive di uso e di distribuzione alla base di ciascun progetto. Gli sviluppatori di eScriptorium, infatti, hanno deciso di seguire una filosofia totalmente Open Access, elaborando un'applicazione web che si appoggia all'infrastruttura di Kraken e permette l'esportazione in locale dei modelli HTR creati dall'utente¹³. Transkribus, al contrario, si presenta come un'infrastruttura che segue un'ottica più commerciale, perché richiede l'ottenimento di crediti per il riconoscimento, una volta esauriti i crediti mensili gratuiti a disposizione dell'utente, e permette la condivisione dei modelli solamente all'interno dello spazio cloud della piattaforma, senza la possibilità di un'esportazione in locale.

Per questa ragione, si è sviluppato negli ultimi anni un intenso dibattito intorno a Transkribus e il suo impiego all'interno di progetti universitari, con un ampio numero di critiche provenienti dai sostenitori dell'Open Access e della libera circolazione delle risorse digitali. A ben vedere, però, bisogna considerare una differenza sostanziale alla base dell'utilizzo delle due piattaforme, soprattutto per quanto ri-

¹⁰ Comunemente la distanza tra due testi si denomina *edit distance* o distanza di Lavenstein: il dato si ottiene calcolando il numero di operazioni (aggiunte, soppressioni, sostituzioni) che consentono di passare da un testo A ad un testo B.

¹¹ L'applicazione è disponibile in due versioni: l'*expert client*, scaricabile alla seguente pagina web <<https://readcoop.eu/transkribus/>>, e l'applicazione web, accessibile dalla medesima pagina. Per maggiori informazioni relative a Transkribus si veda Mühlberger et al. 2019.

¹² Per maggiori informazioni su eScriptorium si consulti Stokes et al. 2021. La documentazione relativa all'installazione e all'utilizzo della piattaforma si trova disponibile al seguente link: <<https://escriptorium.readthedocs.io/en/latest/>>.

¹³ Pinche 2023; Chagué-Clérice 2020-23.

guarda l'installazione, lo sviluppo e la risoluzione dei problemi lato utente. Di fatto, eScriptorium è un software, non un servizio, pertanto non è accessibile da un unico sito web, ma da distinti server alloggiati presso le istituzioni che decidono volontariamente di installare l'applicazione e metterla a disposizione di utenti esterni¹⁴. Questo comporta che l'uso dell'applicazione sia totalmente libero e sostenibile, ma d'altro canto impone delle limitazioni: nel caso di installazione locale, l'utente necessita di competenze informatiche avanzate e di una macchina che possa sostenere un'elevata potenza di calcolo; se invece si decide di appoggiarsi a server già attivi, inevitabilmente si incorrerà in problematiche specifiche del server, per esempio continui aggiornamenti o paralisi di alcuni comandi, a cui spesso si accompagna un supporto tecnico meno efficiente, perché legato alla disponibilità dei gestori del server.

Gli sviluppatori di Transkribus, invece, hanno preferito sviluppare un sistema cloud di appoggio alla piattaforma, attualmente ospitato nei server dell'Università di Innsbruck. Questo comporta che l'accesso al software non richieda all'utente particolari competenze informatiche o tecnologiche, né vi sia la necessità di un supporto istituzionale per l'installazione. La conseguenza diretta di queste scelte è che Transkribus ha visto una più estesa diffusione a partire dal 2018, con un sostanzioso allargamento del bacino di utenti registrati e un incremento del numero di istituzioni partner che partecipano al progetto¹⁵. Poiché i sistemi di *machine learning* alla base di queste piattaforme progrediscono con la quantità di materiali processati, questa espansione della comunità di utenti di Transkribus corrisponde a una maggior efficienza dell'applicazione, che attualmente assicura risultati di segmentazione e trascrizione nettamente migliori rispetto a quelli offerti da eScriptorium. Inoltre, la semplicità di installazione e utilizzo di Transkribus corrisponde a una maggiore distribuzione della piattaforma in contesti accademici, dove gli utenti spesso non hanno esperienza con linea di comando e altre specificità tecniche.

Le piattaforme di HTR come Transkribus ed eScriptorium consentono, come si diceva, di creare un modello di riconoscimento adatto all'interpretazione del testo contenuto in documenti a stampa e manoscritti. In prima battuta, questo modello può basarsi su di un unico documento, del quale l'utente trascrive manualmente un numero determinato di pagine all'interno della piattaforma, con la volontà poi di lanciare il modello HTR creato sulle restanti pagine. In tal caso, si parla di modello HTR individuale, vale a dire di uno strumento specificatamente concepito per l'interpretazione di un singolo testo. Da una parte, questa operazione assicura

¹⁴ Stokes 2021.

¹⁵ Nel momento in cui scrivo la comunità di READ Coop vanta all'incirca 160 membri provenienti da più di 30 Paesi diversi. Tra i membri si contano più di 70 istituzioni partner, tra cui la British Library, archivi nazionali come quelli di Zurigo, Lussemburgo, Norvegia, e varie università europee ed extra-europee. La lista è disponibile al seguente indirizzo: <<https://readcoop.org/members>>.

alcuni vantaggi, per esempio un maggior controllo del flusso di trascrizione automatica nel rispetto dei criteri di trascrizione stabiliti, dal quale deriva la possibilità di prevedere in che punti del testo la macchina potrebbe incontrare maggiori difficoltà. D'altro canto, il limite di un modello individuale è quello di essere efficace solo nel contesto di opere di una certa lunghezza: in presenza di testi brevi, infatti, l'utente non dispone di materiale sufficiente per creare una *ground truth* adeguata. Pertanto, si può con ragione affermare che al momento i modelli individuali sono particolarmente utili nel contesto di testi manoscritti molto estesi, in cui i problemi derivati dalle variazioni interne della grafia, che generalmente complicano l'intero processo di riconoscimento, possono essere superati con la creazione di un imponente dataset di addestramento.

In risposta alla necessità di diffondere l'uso dei sistemi di HTR ad opere di breve estensione, come relazioni, frammenti, articoli di giornale ecc., è stata recentemente introdotta la possibilità di creare modelli misti (*general models*), ossia modelli basati su diversi documenti che vengono raccolti in un unico set di dati e costituiscono la base dell'allenamento¹⁶. I modelli misti possono essere di tipo diverso, a seconda delle opere che compongono il dataset. Nel caso della trascrizione automatica di documenti a stampa, spesso la scelta è quella di includere in un unico modello misto diverse realizzazioni grafiche di uno stesso set di caratteri, per esempio comprendendo opere affini per tipologia o pubblicate da una stessa stamperia in un lasso di tempo determinato. Viceversa, per il riconoscimento di testi manoscritti si preferisce ricorrere ad opere di uno stesso autore, per fare in modo che l'addestramento tenga conto delle diversità della grafia che, benché provenga da una stessa mano, dipende dalle condizioni di scrittura, dal tempo di composizione e dai supporti impiegati. In entrambe le situazioni, inoltre, si consiglia di fondare il modello su opere di diversa provenienza, presupponendo che la varietà interna derivata da tecniche di acquisizione e preservazione distinte estenda le possibilità del modello.

Recentemente, la creazione di modelli misti molto estesi sta attirando l'attenzione della comunità scientifica, anche perché gli stessi modelli misti possono costituire la base di un allenamento personalizzato. In tal caso si parla di *fine tuning*, vale a dire di perfezionamento di un modello misto esteso, che viene usato come *base model* nella creazione di un modello individualizzato che include specifici materiali di *training* trascritti manualmente dall'utente. Questa operazione è largamente praticata, poiché limita la necessità di creare dataset *ex novo* e incentiva il riuso di materiali già processati, ma va sempre sottoposta a un controllo da parte dell'utente perché può comportare una scarsa sistematicità dei risultati di trascrizione, che derivano dall'impiego di un modello HTR di base in parte sconosciuto. Per tale ragione, la corretta descrizione dei dataset che si sono impiegati per la

¹⁶ Hodel et al. 2021.

creazione di modelli misti comincia a rivestire un'importanza sempre maggiore, poiché solamente chiarendo le caratteristiche del modello e i criteri di trascrizione utilizzati si assicura che quegli stessi materiali possano essere utili ad altri utenti. I progetti OCR-D¹⁷ e HTRUnited¹⁸, che sono i principali canali di distribuzione di dataset per la trascrizione automatica, insistono proprio sulla definizione dettagliata dei requisiti di ciascun modello, fomentando la libera circolazione di materiali di training presso altri progetti.

Il Progetto Mambrino: esperienze e prospettive nel campo della trascrizione automatica

Il Progetto Mambrino è un gruppo di ricerca dell'Università di Verona diretto da Anna Bognolo e Stefano Neri che si occupa dello studio dei *libros de caballerías*, romanzi che si diffusero per più di un secolo in tutta la Penisola Iberica a partire dalla pubblicazione nel 1508 dell'*Amadís de Gaula* di Garci Rodríguez de Montalvo. Il genere ebbe un enorme successo in Spagna, con la pubblicazione di più di 60 titoli e diverse riedizioni dei libri appartenenti ai cicli più famosi come quelli di Amadís e Palmerín, sino a varcare i confini nazionali e venire distribuiti in tutta Europa, in Italia, Francia, Inghilterra, Germania, convertendosi in uno dei primi fenomeni di letteratura di massa. In Italia questi romanzi circolarono in un primo momento in lingua originale, con epicentro di produzione prima a Roma e poi a Venezia, grazie alle imprese editoriali dei Nicolini da Sabbio e Giovanni Battista Pederzano¹⁹; ma ben presto, a causa del successo che riscontrava il genere, si iniziarono a stampare traduzioni e continuazioni originali, le principali ad opera di Mambrino Roseo da Fabriano, che tra il 1550 e il 1580 operava in relazione con l'editore Michele Tramezzino, un importante stampatore veneziano²⁰.

Al centro degli interessi del Progetto Mambrino c'è la volontà di riportare alla luce questo genere romanzesco, che tanta fortuna ebbe in Italia durante il Rinascimento, per mezzo dello studio delle opere e il censimento di edizioni ed esemplari conservati. Per rispondere a questa necessità, uno degli obiettivi principali del progetto è quello di pubblicare un'edizione moderna dei testi, realizzata seguendo criteri rigorosi e invalsi nella disciplina. Non si tratta però di un'operazione semplice, vista l'estensione del corpus e delle opere che lo compongono: al riguardo, si consideri che il solo Ciclo di Amadis di Gaula italiano si compone di 21 libri e ciascuno di essi supera le 400 carte²¹.

¹⁷ <https://ocr-d.de/en/>.

¹⁸ <https://htr-united.github.io/>.

¹⁹ Bognolo 2017.

²⁰ Bognolo 2012; Bognolo-Cara-Neri 2013.

²¹ Per uno studio bibliografico delle traduzioni e continuazioni italiane dei *libros de caballerías* si consulti la sezione "Spagnole Romanzerie" del sito web del progetto: <<https://www.mambrino.it/it/spagnole-romanzerie>>.

A partire dal 2016, grazie all'esperienza maturata dai membri del progetto nel campo delle Digital Humanities, ha preso corpo l'idea di realizzare una Biblioteca Digitale dei romanzi cavallereschi italiani di derivazione spagnola, per mezzo della creazione di edizioni scientifiche digitali dei volumi che compongono i cicli di Amadis e Palmerino, le quali, integrate a database bibliografici e semantici, consentono lo studio del contesto editoriale e degli aspetti narrativi del genere²². Ai fini di realizzare un'edizione, in quella circostanza sono state esplorate per la prima volta le possibilità di applicare strumenti di trascrizione automatica per ottenere il testo grezzo delle opere in formato elettronico, poi editabile secondo i criteri stabiliti dal progetto²³.

Dopo le iniziali sperimentazioni con sistemi di OCR come ABBYYFineReader²⁴, che non assicuravano risultati affidabili se applicati a testi in caratteri corsivi, per generare le trascrizioni delle opere si è scelto di impiegare la piattaforma Transkribus (READ Coop), facendo affidamento sulle migliori prestazioni dei sistemi di HTR nell'ambito di testi a stampa antichi. I risultati furono sin da subito impressionanti, con la creazione di modelli individuali per ciascuna opera che si avvicinavano al 99% di precisione, indipendentemente dallo stato di conservazione e dalla qualità delle digitalizzazioni²⁵. I criteri di trascrizione scelti per la creazione del dataset di ogni modello individuale implicavano una modernizzazione del testo, con lo scioglimento delle abbreviazioni e l'interpretazione di caratteri speciali del corsivo, come le legature e il segno tironiano, che complicavano l'applicazione di sistemi di OCR tradizionali. Al contrario, quegli stessi elementi grafici non costituivano un particolare problema per un sistema di HTR come Transkribus, che offriva risultati di trascrizione affidabili, generando testi adatti sia per la creazione di edizioni che per lo studio con strumenti computazionali, per esempio analisi quantitative di stilometria²⁶.

Mentre si proseguiva con la trascrizione automatica e la posteriore correzione dei testi ottenuti con l'intenzione di popolare la Biblioteca Digitale, l'esperienza maturata ispirò nuovi sviluppi nel campo del riconoscimento di testi a stampa antichi. Nel 2021, grazie al sostegno dei gruppi di ricerca spagnoli BIDISO (Biblioteca Digital Siglo de Oro²⁷) e COMEDIC (Catálogo de obras medievales impresas en castellano²⁸), fu possibile avviare un progetto collaborativo che condusse alla creazione di due modelli di HTR per l'interpretazione di testi a stampa iberici in caratteri gotici e romani²⁹. In quell'occasione, si decise per la prima volta di lavorare alla

²² Bognolo-Bazzaco 2019; 2024.

²³ Bazzaco 2018.

²⁴ Mancinelli 2016.

²⁵ Bazzaco 2018, p. 268.

²⁶ Bazzaco 2021.

²⁷ <https://www.bidiso.es/>.

²⁸ <https://comedic.unizar.es/?es>.

²⁹ Bazzaco et al. 2022.

creazione di modelli misti, che integrassero la trascrizione manuale di una porzione di opere che differivano per tipologia editoriale, genere e qualità delle fotoriproduzioni. Guidati dalla volontà che i modelli di HTR sviluppati fossero poi esportabili ad altri contesti di ricerca, per la costituzione della *ground truth* si preferì realizzare una trascrizione manuale conservativa, che riproducesse fedelmente i segni grafici presenti nelle fonti, ad eccezione della “s lunga”, convertita in “s semplice” e del segno tironiano, trascritto come “&”. Le abbreviazioni anche in questo caso si sciolsero, ai fini di ottenere testi finali più facilmente interpretabili e processabili da parte dei collaboratori del progetto³⁰. Inoltre, i modelli si diffusero pubblicamente all’interno della piattaforma Transkribus e furono costantemente implementati, con l’aggiunta di nuove trascrizioni realizzate da altri studiosi interessati, che permisero un ampliamento del dataset di addestramento e, corrispondentemente, un sensibile incremento delle prestazioni di ciascun modello³¹.

La qualità dei risultati ottenuti dimostrò l’alto grado di affidabilità dei sistemi di HTR misti in relazione con i documenti a stampa antichi, gettando le basi per lo sviluppo di un modello che consentisse l’interpretazione generalizzata di testi in caratteri corsivi, utile non solo per le opere al centro degli interessi del Progetto Mambrino, ma esportabile anche ad altri contesti di ricerca.

Verso un modello di HTR per il corsivo del XVI secolo

Nel 2023, alla luce dell’esperienza maturata dal Progetto Mambrino nel campo della trascrizione automatica, ha preso avvio lo sviluppo del modello di HTR misto *Italics_VeniceXVIc*³².

1. *Selezione dei materiali di addestramento.* La prima fase per la creazione del dataset all’interno della piattaforma Transkribus corrisponde alla scelta delle opere che andranno a costituire la base per l’addestramento della macchina. Per eseguire questa operazione, sono stati selezionati alcuni romanzi cavallereschi italiani di interesse del progetto, secondo criteri di variabilità dei caratteri tipografici e diversità dei processi di acquisizione delle immagini (Fig. 1).

³⁰ Bazzaco et al. 2022, p. 95.

³¹ Si segnalano di seguito le principali caratteristiche dei due modelli:

- Modello *SpanishGothic_XV-XVI_extended* (vers. 1.2.0). Stefano Bazzaco (coord.), Federica Zoppi, Giada Blasut, Nuria Aranda García, Ángela Torralba Ruberte, Ana-Milagros Jiménez Ruiz, Pedro Monteiro, José Manuel Fradejas, Eduardo Camero Santos, Laura Lecina Nogués, Almudena Izquierdo Andreu. Dataset disponibile all’indirizzo: <<https://doi.org/10.5281/zenodo.4888926>>.
- Modello *SpanishRedonda_XVI-XVII_extended* (vers. 1.2.0). Stefano Bazzaco (coord.), Gaetano Lalomia, Daniela Santonocito, Manuel Garrobo Peral, Mónica Martín Molares, Carlota Cristina Fernández Travieso, Giulia Tomasi, Alessia Fichera, Soledad Castaño Santos, Almudena Izquierdo Andreu. Dataset disponibile all’indirizzo: <<https://doi.org/10.5281/zenodo.4889217>>.

³² Lo sviluppo del modello *Italics_VeniceXVIc* è stato realizzato da chi scrive in collaborazione con Giulia Lucchesi (Università di Verona), che ringrazio.

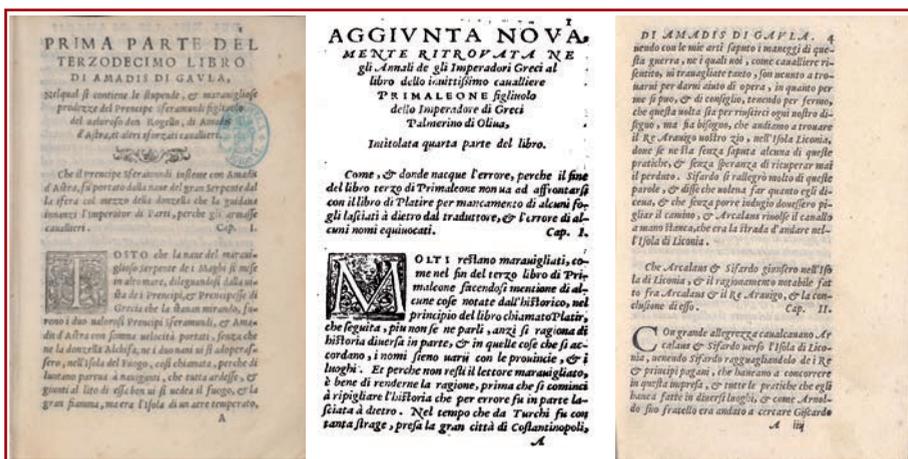


Figura 1. Esempi di materiali inclusi nel corpus, tratti da Sferamundi di Grecia. Prima parte; La quarta parte di Primaleone; Aggiunta al quarto libro di Amadis di Gaula (in ordine da sinistra a destra).

Si offre di seguito la lista delle opere incluse nel modello, seguendo l'ordine cronologico di pubblicazione di ciascun esemplare:

Autore/traduttore e Opera	Data	Editore	Provenienza digitalizzazioni
Mambrino Roseo da Fabriano, <i>Flortir</i>	1558	Michele Tramezzino	Praha, Národní knihovna České republiky, 9 J 000225
Mambrino Roseo da Fabriano, <i>Sferamundi di Grecia. Prima parte</i>	1558	Michele Tramezzino	Madrid, Biblioteca Nacional de España, 5-4978
Francesco Portonaris, <i>Palmerino d'Inghilterra</i>	1558	Francesco Portonaris	Wien, Österreichische Nationalbibliothek, 8.J.33.(Vol.1)
Mambrino Roseo da Fabriano, <i>Sferamundi di Grecia. Seconda parte</i>	1560	Michele Tramezzino	Madrid, Biblioteca Nacional de España, 5-4978
Mambrino Roseo da Fabriano, <i>La quarta parte di Primaleone</i>	1560	Michele Tramezzino	Wien, Österreichische Nationalbibliothek, O.M.40.(Vol.2)
Mambrino Roseo da Fabriano, <i>Aggiunta al quarto libro di Amadis di Gaula</i>	1563	Michele Tramezzino	Santiago de Compostela, Biblioteca Universitaria, 13996
Mambrino Roseo da Fabriano, <i>Aggiunta al Florisello (Le prodezze di don Florarlano)</i>	1564	Michele Tramezzino	Verona, Biblioteca Civica, Cinq. E 350-12
Mambrino Roseo da Fabriano, <i>Aggiunta a Rogello di Grecia</i>	1564	Michele Tramezzino	Verona, Biblioteca Civica, Cinq. 350-15

Le digitalizzazioni di queste opere, che erano già state importate all'interno della piattaforma in occasione della creazione di modelli individuali specifici, sono state raccolte in un'unica collezione all'interno della piattaforma: di ciascuna di esse si sono selezionate 20 pagine interne corrispondenti al testo in prosa, in modo da confezionare un dataset sufficientemente equilibrato, in cui nessun testo prevaricasse sugli altri in termini di quantità.

2. *Definizione dei criteri di trascrizione.* Una volta selezionati i materiali di addestramento, si è passati alla definizione dei criteri di trascrizione. Come già anticipato, questa operazione richiede un'attenta valutazione da parte dello studioso, poiché le scelte di trascrizione incidono in maniera determinante sui risultati finali dell'allenamento del modello. Come stabilito in Heigl, «the machine must learn from a transcription that is as accurate as possible so that it can later reproduce exactly what is written on the sheet» (2019), pertanto è preferibile una trascrizione diplomatica. Questo suggerimento, che non genera problemi, per esempio, nel caso della trascrizione del segno tironiano come “&” o nel mantenimento della distinzione “u/v”, si complica nel caso delle abbreviature, che possono corrispondere a più di un carattere nel testo trascritto. L'utente al riguardo ha a disposizione tre diverse possibilità: sviluppare l'abbreviazione, per rendere il testo finale più comodamente ricercabile; mantenere il carattere speciale in corrispondenza dell'abbreviazione, possibilmente facendo ricorso a un set di caratteri standard, come “Latin Extended-D” di Unicode; trascrivere il carattere speciale come carattere semplice, aggiungendo però una metadatarazione di tipo sintattico che indichi che siamo in presenza di un'abbreviazione, per esempio aggiungendo il tag <abbr>³³. Quest'ultima soluzione, come si può immaginare, richiede un maggior tempo di produzione rispetto alle precedenti; per questo è una pratica poco seguita.

La decisione sui criteri è sempre una fase delicata, che risponde a delle necessità concrete e che comporta dei vantaggi, ma anche, nella maggior parte dei casi, una perdita di informazione. Per esempio, nel caso si confezioni una *ground truth* che presenta una modernizzazione del testo fonte, la trascrizione che si potrebbe ottenere applicando il modello non terrebbe conto della presenza di eventuali “s lunghe” (che verrebbero a uniformarsi con le “s semplici”). In caso contrario, una trascrizione diplomatica potrebbe complicare una normalizzazione automatica dei caratteri, per esempio limitando la regolarizzazione di “u/v” secondo l'uso attuale. Nel nostro caso, per lo sviluppo di un primo modello di riconoscimento per i caratteri corsivi si è scelto di adottare dei criteri di trascrizione conservativi, secondo una prospettiva di preservazione di segni grafici speciali e soluzioni tipografiche che potrebbero risultare di interesse per settori di studio specifici, per esempio la

³³ Le medesime indicazioni si ritrovano in Heigl 2019 e nelle Guidelines di Transkribus: <<https://help.transkribus.org/data-preparation>>.

paleografia o la bibliografia materiale dei testi a stampa³⁴. In futuro si prevede però la pubblicazione di un secondo modello di riconoscimento che consenta la produzione di un testo modernizzato, in modo da favorire il processamento con strumenti di analisi quantitativa.

3. *Segmentazione e trascrizione manuale*. L'operazione di *Layout Analysis* è stata avviata all'interno della piattaforma per mezzo di due comandi principali: in primo luogo, attraverso la funzione "Printed Block Detection" sono state individuate le regioni di testo; successivamente, sono state individuate in modo automatico le linee di testo, avendo cura di interrompere ciascuna riga in corrispondenza del perimetro delle regioni. Infine, si è provveduto alla trascrizione manuale delle porzioni di ogni opera, secondo i criteri prestabiliti.

4. *Creazione del modello*. L'addestramento del modello *Italics_VeniceXVIc* è stato realizzato selezionando dalla *ground truth* il materiale di addestramento e segnalando una corrispondente porzione di testo, pari al 10% dell'intero dataset, come materiale di validazione. Nello specifico, sono state impiegate 4.848 linee di testo trascritto manualmente, corrispondenti a 40.982 parole processate.

Per l'allenamento si è scelto un numero massimo di *epochs* pari a 350, con una funzione di *early stopping* fissata a 20 *epochs*. Ciò significa che, valutando la coincidenza tra le curve di apprendimento di *training* e *validation set*, e quindi l'inefficacia di ulteriori fasi di addestramento, l'applicazione poteva interrompere l'allenamento dopo il ventesimo periodo: nel caso concreto del modello creato, l'allenamento si è interrotto all'ottantacinquesima *epoch* (Fig. 2).

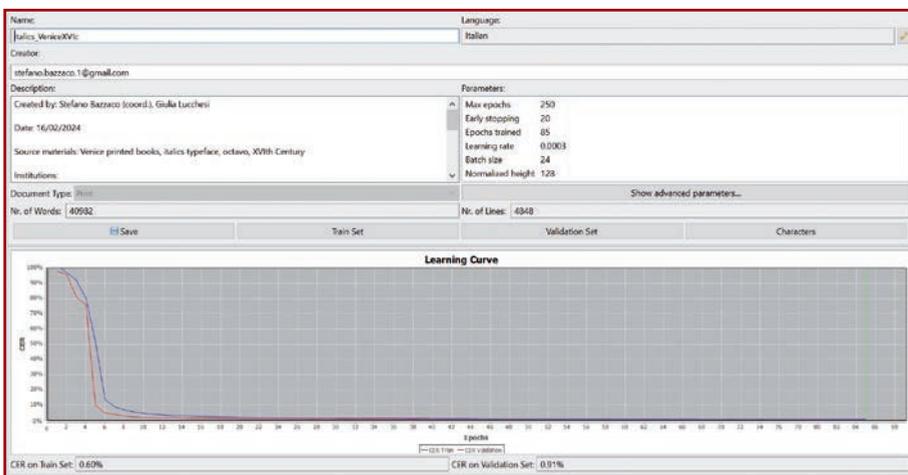


Figura 2. Dettagli e curve di apprendimento del modello *Italics_VeniceXVIc*

³⁴ Per esempio, si sono conservati caratteri speciali presenti nella fonte come $\grave{}$ (per la 's lunga') o p (per le abbreviazioni di "per", "pro", "pre"). Una visualizzazione dettagliata dei criteri è consultabile alla pagina: <https://github.com/stefanobazzaco/HTR-model-italics_VeniceXVIc>.

5. *Valutazione e Fine Tuning*. I risultati dell'addestramento, come si è detto, vengono trasmessi secondo un indice chiamato CER, che nel contesto che ci riguarda arriva al 0.91%. L'indice di errore, però, riguarda materiali simili a quelli processati dall'applicazione, vale a dire quelli inclusi nel set di validazione; per questa ragione, è necessario testare la concreta efficienza del modello su nuovi materiali sconosciuti alla macchina. Questa fase di valutazione concreta dei risultati ottenibili, nel contesto del modello creato, è stata particolarmente efficace, perché il riconoscimento su altri testi cavallereschi affini ha rivelato una difficoltà di interpretazione di alcuni elementi presenti sulla pagina, per esempio i titoli e le intestazioni in lettera maiuscola.

Per risolvere queste limitazioni, è possibile realizzare un'operazione di *fine tuning*, vale a dire di perfezionamento del modello. Questa fase potrebbe prevedere l'integrazione di più modelli o l'aggiunta di nuovi materiali al dataset iniziale, opportunamente scelti in relazione con le sezioni del testo dove il riconoscimento automatico rivela maggiori debolezze. Nell'ambito del modello qui descritto è stato impiegato un altro modello di HTR pubblico, chiamato *Transkribus_PrintM1*, come *base model*³⁵. L'operazione si è rivelata estremamente utile, perché ha permesso di ridurre ulteriormente il CER, sceso al 0.70%, e ha assicurato risultati più affidabili anche in corrispondenza dei caratteri che non venivano interpretati correttamente in precedenza (Fig. 3).



Figura 3. Perfezionamento dei risultati grazie all'operazione di Fine Tuning (al centro, modello *Italics_VeniceXVIc*; a destra, modello *Italics_VeniceXVIc+basemodel*)

³⁵ Una descrizione delle caratteristiche di *Transkribus_PrintM1* può essere consultata nella sezione "Public Models" di READCoop: <<https://readcoop.eu/it/model/transkribus-print-multi-language-dutch-german-english-finnish-french-swedish-etc/>>.

Come si può osservare, il modello perfezionato non risente dei criteri di trascrizione del modello di supporto *Transkribus_PrintM1*, che prevale solamente in presenza di punti particolarmente critici per il modello *Italics_VeniceXVIc*. Per esempio, si noti la preservazione della “β”, presente nel modello perfezionato ma trascritta come “s semplice” nel *base model*.

6. *Distribuzione e sviluppi futuri*. Forti dell’esperienza maturata con i testi a stampa antichi di area iberica, nell’orbita del Progetto Mambrino si sta valutando la possibilità di implementare il modello attraverso la collaborazione di altri studiosi, che potranno procedere con l’elaborazione di nuove trascrizioni manuali di porzioni di testi simili a quelli già inclusi ed integrarle ai materiali di addestramento. In tal senso, si prevede la distribuzione di una prima versione del modello *Italics_VeniceXVIc* nel mese di giugno 2024, suscettibile di future integrazioni. La *ground truth* di base del modello è attualmente disponibile nella piattaforma Zenodo³⁶ e contiene i materiali estratti dalla piattaforma Transkribus: le immagini originali, i corrispondenti file nei formati standard ALTO XML e PAGE XML, nonché i file in testo piano (TXT). La costituzione del dataset pertanto rispetta le indicazioni di distribuzione e riutilizzo dei dati suggerite dal progetto HTR-United³⁷ e, indipendentemente dai futuri aggiornamenti, potrà essere facilmente individuabile a partire da un unico DOI (Digital Object Identifier), che restituirà all’utente l’ultima versione pubblicata.

In seguito, per favorire la creazione di uno strumento che consenta lo scioglimento automatico di abbreviazioni e simboli speciali, adattando quindi le trascrizioni risultanti a nuovi scenari di impiego, si contempla la possibilità di modificare la *ground truth* di partenza secondo nuovi criteri di trascrizione e di procedere all’allenamento di un secondo modello chiamato *Italics_VeniceXVIc_modernized*. Entrambi i modelli saranno resi disponibili nella piattaforma Transkribus (sezione “Public Models”) e le rispettive *ground truth* verranno successivamente impiegate nella creazione di modelli affini con la piattaforma eScriptorium, in modo da estendere il loro utilizzo presso una comunità scientifica più ampia.

Conclusioni

Come già chiarito, lo studio che si propone in questo articolo costituisce un primo passo verso la creazione di un modello di HTR che consenta il riconoscimento automatico e generalizzato di testi a stampa in carattere corsivo. L’argomento risulta di estrema attualità, visti i recenti avanzamenti nel campo della trascrizione automatica che, di pari passo con lo sviluppo tecnologico, sta sperimentando un rinnovato interesse nel contesto accademico e archivistico; basti pensare ai recenti pro-

³⁶ <<https://zenodo.org/records/10674282>>, DOI: 10.5281/zenodo.10674282.

³⁷ Chagué-Clérice 2020-23.

getti di area italiana che hanno preso avvio negli ultimi anni e che propongono un'integrazione dei sistemi di HTR alla ricerca storica e letteraria³⁸.

Nel gruppo di ricerca Progetto Mambrino, in quanto ispanisti interessati alle relazioni tra la letteratura spagnola e quella italiana, ci si è avvicinati a questo campo della ricerca guidati dalla necessità di trascrivere dei romanzi cavallereschi molto estesi, ma la volontà è quella di condividere la nostra esperienza, che risulterebbe molto utile anche a diversi campi dell'italianistica, con cui ci auspichiamo di collaborare. La proposta di lavoro che si descrive in queste pagine, infatti, mira a incentivare lo sviluppo e impiego di sistemi di HTR nel contesto di opere a stampa cavalleresche della metà del Cinquecento, ma la sperimentazione con questi strumenti si può estendere anche ad altri set di caratteri³⁹. Nel nostro caso, i sistemi di HTR, già allo stato attuale, assicurano con il corsivo risultati di trascrizione altamente affidabili, con l'ottenimento di testi elettronici che presentano un margine di errore inferiore all'1%. Raggiungere un tale margine di precisione rappresenta un avanzamento notevole, perché gli stessi testi possono poi essere processati con strumenti computazionali, essere impiegati nello sviluppo di edizioni scientifiche digitali e amplificare le possibilità di collazione nel caso di tradizioni testuali complesse, suggerendo nuovi percorsi di integrazione tra le diverse aree di studio che compongono l'estesa galassia delle Digital Humanities.

³⁸ Zappulli-Iorio 2018; Schwarz-Ricci 2022; Spina 2022; Malatesta 2023.

³⁹ Nel momento in cui scrivo, siamo coinvolti in due progetti italiani che prevedono lo sviluppo di modelli HTR per incunaboli e testi a stampa:

- Progetto PRIN 2022 *spaNice: Spanish cultural models in Early Modern Venice (the development and circulation of Spanish literature and language in 16th-17th century Italy)*, dir. Federica Zoppi (prot. 202297ATKC);
- Progetto PRIN 2022 *Chivalric Spaces. A digital landscape of texts, stories and narrative motifs of printed popular chivalric literature in Italy and Spain*, direttrice Annalisa Perrotta (prot. 2022HT3XYP).

The contribution aims to describe the workflow that led to the production of a Handwritten Text Recognition (HTR) model for the automatic transcription of sixteenth century Venetian texts printed in italics. In the first part, the scope of the study is defined, with particular attention to the state of the art and recent developments in the field of HTR regarding complex scripts, namely historical printed texts and manuscripts that, due to their characteristics, hinder the application of traditional Optical Character Recognition (OCR) systems. The second part outlines the main phases of the training process for the creation of the Italics_VeniceXVIs model, which represents a first step towards the interpretation of chivalric texts in italics of interest to the Mambrino Project at the University of Verona. Finally, the main characteristics of the model are identified, and, with a view to accessibility and reusability, future steps of the project are highlighted, suggesting possible implications of the research carried out in relation to other fields of study.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

RIFERIMENTI BIBLIOGRAFICI

- Bazzaco 2018 Stefano Bazzaco. *El Proyecto Mambrino y las tecnologías OCR: estado de la cuestión*. «Historias Fingidas», 6 (2018), p. 257-272. <<http://dx.doi.org/10.13136/2284-2667/89>>.
- Bazzaco 2021 Stefano Bazzaco. *Experimentos de estilometría en el ámbito de los libros de caballerías. El caso de atribución de un original italiano: Il terzo libro di Palmerino d'Inghilterra (Portonari, 1559)*. In: "Prenga xascú ço qui millor li és de mon dit". *Creació, recepció i representació de la literatura medieval*, coord. Meritxell Simó, ed. G. Avenoza, A. Contreras, G. Sabaté, L. Soriano. Cilengua: San Millán de la Cogolla, 2021, p. 149-166.
- Bazzaco et al. 2022 Stefano Bazzaco — Mónica Martín Molares — Ana Milagros Jiménez Ruiz — Ángela Torralba Ruberte. *Sistemas de reconocimiento de textos e impresos hispánicos de la Edad Moderna. La creación de unos modelos de HTR para la transcripción automatizada de documentos en gótica y redonda (s. XV-XVII)*. «Historias Fingidas», Número Especial 1 (2022), *Humanidades Digitales y estudios literarios hispánicos*, p. 67-125. <<https://historiasfingidas.dlss.univ.it/article/view/1190>>.
- Bognolo 2012 Anna Bognolo. *El libro español en Venecia en el siglo XVI*. In: *Rumbos del hispanismo en el umbral del Cincuentenario de la AIH*, coord. Patrizia Botta, Aviva Garribba, María Luisa Cerrón Puga, Debora Vaccari, vol. 3 (III. Siglo de Oro (prosa y poesía)). 2012, p. 243-258.
- Bognolo 2017 Anna Bognolo. *Entre Celestinas, novela sentimental y libros de caballerías. La empresa editorial de los Nicolini da Sabbio y Juan Bautista Pederzano en Venecia alrededor de 1530*. In: *Serenísima palabra. Actas del X Congreso de la Asociación Internacional Siglo de Oro (Venecia, 14-18 de julio de 2014)*. Venezia: Edizioni Ca' Foscari, 2014, p. 727-738.
- Bognolo-Bazzaco 2019 Anna Bognolo — Stefano Bazzaco. *Tra Spagna e Italia: per un'edizione digitale del Progetto Mambrino*. «eHumanista/IVITRA», 16 (2019), p. 20-36.
- Bognolo-Bazzaco 2024 Anna Bognolo — Stefano Bazzaco. *Editar libros de caballerías en la era digital: la Biblioteca Digital del Proyecto Mambrino*. In: *Editar el Siglo de Oro en la era digital*, ed. E. Fosalba, S. Allés Torrent. Studia Aurea Monográfica, n. 9, p. 19-48.
- Bognolo-Cara-Neri 2013 Anna Bognolo — Giovanni Cara — Stefano Neri. *Repertorio delle continuazioni italiane ai romanzi cavallereschi spagnoli. Ciclo di Amadis di Gaula*. Roma: Bulzoni Editore, 2013.
- Chagué-Clérice 2020-2023 *HTR-United (2020-2023)*, ed. by Chagué, Alix — Thibault Clérice. HTR-United Catalog. <<https://github.com/HTR-United/htr-united/>>.

- Ciotti 2023 Fabio Ciotti. *Minerva e Il Pappagallo. IA generativa e modelli linguistici nel laboratorio dell'umanista digitale*. «Testo E Senso», (dicembre 2023), n. 26, p. 289-15.
<<https://doi.org/10.58015/2036-2293/671>>.
- Cordell-Smith 2018 Ryan Cordell — David Smith. *Report: A Research Agenda for Historical and Multilingual Optical Character Recognition*. Northeastern University Library.
<<http://hdl.handle.net/2047/D20297452>>.
- Heigl 2019 Elisabeth Heigl. *Transcription Guidelines*. «*Rechtsprechung Im Ostseeraum: Digitization & Handwritten Text Recognition*» (blog), 2019.
<<https://rechtsprechung-im-ostseeraum.archiv.uni-reifswald.de/transcriptionguidelines/>>.
- Hodel et al. 2021 Tobias Hodel — David Schoch — Christa Schneider — Jake Purcell. *General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example*. «*Journal of Open Humanities Data*», 7 (July 2021).
<<https://doi.org/10.5334/johd.46>>.
- Malatesta 2023 Serena Malatesta. *(Proto)modello di trascrizione automatica per i manoscritti danteschi in Littera Textualis (XIV-XV sec.) con Transkribus*. «*Digitalia. Rivista del digitale nei beni culturali*», 18 (2023), n. 2, p. 282–285.
<<https://digitalia.cultura.gov.it/article/view/3016>>.
- Mancinelli 2016 Tiziana Mancinelli. *Early printed edition and OCR techniques: what is the state-of-art? Strategies to be developed from the working-progress Mambrino project work*. «*Historias Fingidas*», (2016), n. 4, p. 255-260.
- Mühlberger et al. 2019 Günter Mühlberger et al. *Transforming scholarship in the archives through Handwritten Text Recognition. Transkribus as a case study*. «*Journal of Documentation - Emerald Publishing*», 75 (2019), n. 5, p. 954-976.
- Pavlopoulos et al. 2022 John Pavlopoulos et al. *Error Correcting HTR'ed Byzantine Text*. «*HTREC 2022*» p. 1-16. <<http://dx.doi.org/10.21203/rs.3.rs-2921088/v1>>.
- Perdiki 2023 Elpida Perdiki. *Preparing Big Manuscript Data for Hierarchical Clustering with Minimal HTR Training*. «*Journal of Data Mining and Digital Humanities. Special Issue: Historical documents and automatic text recognition*», 2023.
<<https://hal.science/hal-03880102v4>>.
- Pinche 2023 Ariane Pinche. *Generic HTR Models for Medieval Manuscripts. The CREMMALab Project*. «*Journal of Data Mining and Digital Humanities. Special Issue: Historical documents and automatic text recognition*», 2023. <<https://hal.science/hal-03837519v3>>.

- Rabus 2019 Achim Rabus. *Recognizing Handwritten Text in Slavic Manuscripts: A Neural-Network Approach Using Transkribus*. «Scripta & e-Scripta», (2019), n. 19, p. 9–32.
- Roncaglia 2023 Gino Roncaglia. *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*. Bari: Laterza, 2023.
- Sayre 1973 Kenneth M. Sayre. *Machine recognition of handwritten words: A project report*. «Pattern Recognition», 5 (1973), n. 3 p. 213-228.
<[https://doi.org/10.1016/0031-3203\(73\)90044-7](https://doi.org/10.1016/0031-3203(73)90044-7)>.
- Schwarz-Ricci 2022 Vera Isabel Schwarz-Ricci. *Handwritten Text Recognition per registri notariili (secc. XV-XVI): una sperimentazione*. «Umanistica Digitale», (2022), n. 13, p. 171-181.
<<https://doi.org/10.6092/issn.2532-8816/14926>>.
- Spina 2022 Salvatore Spina. *Historical Network Analysis and HTR tools for a digital methodological historical approach to the Biscari Archive of Catania*. «Umanistica Digitale», (2022), n. 14, p. 163–181.
<<https://doi.org/10.6092/issn.2532-8816/15159>>.
- Stokes et al. 2021 Peter A. Stokes — Benjamin Kiessling — Daniel Stökl Ben Ezra — Robin Tissot — El Hassane Gargem. *The EScriptorium VRE for Manuscript Cultures*. «Classics@ Journal, Ancient Manuscripts and Virtual Research Environments» ed. C. Clivaz e G. V. Allen, 18 (2021).
<<https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/>>.
- Stokes-Kiessling 2024 Peter A. Stokes — Benjamin Kiessling. *Sharing Data for Handwritten Text Recognition (HTR)*. «Digital Humanities in Practice», In corso di stampa. fihal-04444641
<<https://hal.science/hal-04444641/document>>.
- Terras 2022 Melissa Terras. *Chapter 7: Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription*. In: *Archives, Access and Artificial Intelligence. Working with Born-Digital and Digitized Archival Collections*, ed. by L. Jaillant. Bielefeld University Press, 2022, p. 179-204.
<<https://doi.org/10.14361/9783839455845-008>>.
- Vinciarelli 2003 Alessandro Vinciarelli. *Offline Cursive Handwriting: From Word To Text Recognition*. IDIAP Research Record 03-24, 2003.
<<https://infoscience.epfl.ch/record/82879>>.
- Zappulli-Iorio 2018 Andrea Zappulli — Sabrina Iorio. *La digitalizzazione dell'Archivio Storico del Banco di Napoli*. «DigItalia. Rivista del digitale nei beni culturali», 13 (2018), n. 2 p. 46-51.
<<https://digitalia.cultura.gov.it/article/view/2169>>.

Progetti

I libri Sapienza parlano. Un progetto per l'accessibilità dei libri della Sapienza alle persone con disabilità visive e DSA

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00095

**Laura Armiero — Gianfranco Crupi — Angela Di Iorio — Loredana Di Lucchio
Agnese Galeffi — Alessandra Gulotta — Roberto Raieli — Valentina Rovacchi**
Sapienza Università di Roma

L'articolo illustra il progetto di terza missione della Sapienza Università di Roma I libri Sapienza parlano, indicandone genesi, finalità, fasi di sviluppo e attività realizzate. Il progetto intende fornire alle persone con disabilità visive e DSA, anche non appartenenti alla comunità Sapienza, versioni in formati accessibili dei libri presenti nelle biblioteche di ateneo. Sapienza Università di Roma ha già ottenuto l'autorizzazione a effettuare copie accessibili grazie all'eccezione al diritto d'autore in favore di alcune categorie di beneficiari.

1. Introduzione

Il progetto di terza missione *I libri Sapienza parlano*, coordinato dal Sistema bibliotecario Sapienza (SBS)², con la collaborazione di bibliotecarie e bibliotecari afferenti a diverse aree disciplinari, è nato su sollecitazione della delegata della rettrice per le iniziative in favore degli studenti con disabilità e con DSA, professoressa Caterina Lombardo, ed è stato finanziato dall'Ateneo per il periodo 2022-2024³. Il progetto ha inteso rendere concrete alcune opportunità in materia di accessibilità offerte dalla legislazione italiana⁴ ed europea⁵, per consentire ad alcune categorie di beneficiari (non vedenti, ipovedenti, con disabilità percettive o di lettura, e altre disabilità fisiche) la riproduzione di talune tipologie di testi, usufruendo dell'eccezione al diritto d'autore prevista dalla normativa nazionale.

¹ L'introduzione è a cura di Valentina Rovacchi; il par. 2 è dovuto a Roberto Raieli; il par. 3.1 è dovuto ad Agnese Galeffi e Angela Di Iorio; i par. 3.2 e 3.3 sono di Laura Armiero; il par. 3.4 è di Alessandra Gulotta; il par. 4. è stato scritto da Loredana Di Lucchio; l'editing complessivo e le conclusioni sono di Gianfranco Crupi.

² <https://web.uniroma1.it/sbs/>.

³ Tutte le informazioni sul progetto sono disponibili alla pagina: <<http://tinyurl.com/zn6rckm>>. Il progetto è stato presentato presso la Sapienza Università di Roma il 22 novembre 2022 e, successivamente (11 dicembre), nel corso della Fiera nazionale della piccola e media editoria, "Più libri più liberi".

⁴ Ci si riferisce all'art. 71-bis, comma 2-quinquies della legge n. 633/1941 e al d.p.c.m. 6/7/2020.

⁵ Il riferimento è alla Direttiva UE 2017/1564: <<http://tinyurl.com/yk2s3x5n>>.

La direttiva europea 2017/1564 stabilisce che copie in formato accessibile di un'opera o di altro materiale, cui le persone disabili hanno legittimamente accesso, possono essere realizzate da «entità autorizzate», che sono tenute a dichiarare il possesso dei requisiti previsti dalla legge e a rispettare una serie di obblighi. La verifica del possesso dei requisiti e del rispetto degli obblighi è disciplinata in Italia dal d.p.c.m. del 6/7/2020.

Il primo passo del progetto è consistito, dunque, nella richiesta da parte della Sapienza del riconoscimento di entità autorizzata a riprodurre testi in formato accessibile⁶, con la finalità di fornire ai destinatari dell'iniziativa l'opportunità di consultare la stessa varietà di testi di cui possono usufruire coloro che disabilità non hanno, compresi i testi di natura scientifica, saggistica, divulgativa.

Coerentemente alla natura di terza missione del progetto, che prevede l'apertura delle università ai contesti territoriali, attraverso il trasferimento di conoscenze e di competenze, l'offerta dei formati accessibili realizzati si indirizza di conseguenza a una più larga platea di potenziali fruitori, indipendentemente dalla loro appartenenza alla comunità Sapienza.

2. Il contesto

Negli ultimi venti anni circa, con l'uso sempre più massivo del web e delle risorse digitali, nonché dei sistemi di Information retrieval (IR), è cresciuto il numero di studi nel settore Library and information science (LIS) riguardanti il modo in cui gli utenti con disabilità visive cercano di soddisfare le proprie esigenze informative⁷. E tuttavia il mondo delle biblioteche deve ancora acquisire la necessaria consapevolezza circa le specifiche esigenze, strumentali e concettuali, necessarie a questa speciale categoria di utenti, per accedere alle fonti bibliografiche della conoscenza e usare le risorse informative (cataloghi, database, discovery tool, sistemi di IR ecc.). L'acquisizione di questa consapevolezza va inoltre estesa ai servizi bibliotecari, all'accessibilità e usabilità dei sistemi di IR e alla progettazione di idonee risorse bibliografiche cartacee o elettroniche che oggi dipendono sempre più direttamente dalle tecnologie, rese disponibili per facilitare l'approccio ai materiali da leggere⁸. Il soddisfacimento delle necessità delle persone con disabilità visive rappresenta, dunque, un obiettivo fondamentale da raggiungere nella ricerca di ambito LIS, oltre che nella produzione editoriale: a parte la presenza di libri in Braille, visori per l'ingrandimento o libri stampati a caratteri grandi, solo in tempi recenti cominciano a diffondersi libri tattili, audiolibri, formati elettronici leggibili dal

⁶ <https://www.librari.beniculturali.it/it/diritto-dautore/Focus-activity/>.

⁷ Iris Xie – Shengang Wang – Meredith Saba, *Studies on blind and visually impaired users in LIS literature: a review of research methods*, «Library and Information Science Research», 43 (2021), n. 3, <<https://doi.org/10.1016/j.lisr.2021.101109>>.

⁸ Gerd Berget – Andrew MacFarlane, *What is known about the impact of impairments on information seeking and searching?*, «Journal of the Association for Information Science and Technology», 71 (2020), n. 5, p. 596-611, <<https://doi.org/10.1002/asi.24256>>.

computer o lettori di schermate. Questi ultimi, in particolare, marcano la differenza nella ricerca tramite le interfacce di IR, distinguendo tra utenti in grado di vedere bene e coloro che non possono: da questo punto di vista, lo sviluppo delle tecnologie dell'informazione degli ultimi anni sta contribuendo significativamente a superare tale differenziazione⁹, tenuto conto che la preferenza dei lettori con disabilità va ai formati elettronici, che possono essere più facilmente maneggiati autonomamente con le tecnologie¹⁰.

Fino alla *Millenium Declaration* delle Nazioni Unite del 2000 non si faceva cenno a politiche globali che prendessero in seria considerazione l'accesso all'informazione dei disabili visivi¹¹: solo sporadiche iniziative, a partire dagli anni Dieci del secolo attuale, hanno valutato la necessità di formulare raccomandazioni per la creazione di oggetti informativi accessibili per i non vedenti¹².

Un ostacolo alla piena attuazione di politiche di accessibilità al patrimonio bibliografico è senz'altro costituito dalle leggi restrittive sul diritto d'autore che limitano l'accesso alle informazioni per tutte le persone, comprese quelle che vivono con disabilità. Basti pensare che le leggi sul copyright non consentono, in varie parti del mondo, nemmeno la trascrizione o la conversione delle risorse informative in Braille¹³. E sebbene le tecnologie assistive abbiano realizzato una serie di prodotti che consentono di accedere facilmente ai testi digitali, la protezione dei diritti introdotta dall'altra tecnologia, quella dei Digital Right Management (DRM), spesso impedisce o complica il funzionamento di questi¹⁴. Molti editori, fortunatamente, hanno iniziato da tempo a rilasciare versioni *DRM-free* per i disabili, e alcuni non si preoccupano più di impedire la lettura automatica dei PDF scaricabili dai propri siti dopo regolare acquisto, anche di tipo "campus".

Per le biblioteche, le insidie per un accesso all'informazione realmente senza di-

⁹ Iris Xie [et al.], *Orientation tactics and associated factors in the digital library environment: comparison between blind and sighted users*, «Journal of the Association for Information Science and Technology», 72 (2021), p. 995-1010, <<https://doi.org/10.1002/asi.24469>>.

¹⁰ Rebecca M. Majinge – Christine Stilwell, *ICT use in information delivery to people with visual impairments and in wheelchairs in Tanzanian university libraries*, «African Journal of Library, Archives and Information Science», 24 (2014), n. 2, p. 151-159, <<https://www.ajol.info/index.php/ajlais/article/view/135069>>.

¹¹ United Nations, *Disability and the millennium development goals: a review of the MDG process and strategies for inclusion of disability issues in millennium development goal efforts*, New York: United Nations, 2011, <<http://tinyurl.com/47t6hkmf>>.

¹² Come, per esempio, i "World Summit on Information Society (WSIS)": Michael Ashley Stein – Janet E. Lord, *Future prospects for the United Nations Convention on the rights of persons with disabilities*, 2008.

¹³ Hee-Yoon Yoon – Sin-Young Kim, *On the improvement of the Copyright Law of Korea for library services for persons with disabilities*, «Journal of Librarianship and Information Science», 45 (2012), n. 2, p. 140-152, <<https://doi.org/10.1177/0961000612457106>>.

¹⁴ Rekha Devi – Shailendra Kumar, *Digital Right Management and accessibility of libraries electronic resources for blind and visually impaired users: a review*, «DESIDOC Journal of Library & Information Technology», 43 (2023), n. 3, p. 193-201, <<https://doi.org/10.14429/djlit.43.03.18544>>.

stinzioni sono, quindi, di vario genere e riguardano le politiche di accesso, l'inadeguatezza dei progetti, i diritti di produzione di formati accessibili, gli aspetti economici, l'inadeguata progettazione dei siti web e delle interfacce di ricerca, la scarsa conoscenza della problematica, la non preparazione all'assistenza, la mancanza in genere di tecnologie assistive (scanner OCR, lettori di schermate, visori, audiolibri ecc.), l'assenza di percorsi tattili all'interno degli edifici. Inutile aggiungere che le biblioteche sono attori fondamentali nella creazione di politiche di inclusione a partire dall'attuazione di programmi di sensibilizzazione e formazione del personale, alla definizione delle politiche di servizio e alla riorganizzazione dello spazio fisico. Purtroppo, in buona parte del mondo bibliotecario, non c'è una diffusa coscienza della disabilità, non si conoscono le tecnologie e non si è preparati a offrire e promuovere i giusti servizi¹⁵.

Le biblioteche delle università, in particolare, hanno l'obbligo istituzionale – oltre che etico e morale – di garantire a tutte le tipologie di utenti di fruire pienamente del patrimonio bibliografico e delle risorse dell'informazione da esse messe a disposizione, allo scopo di consentire a ognuno, tramite l'istruzione universitaria, di raggiungere i più alti livelli di sviluppo scientifico, professionale, economico e sociale¹⁶.

Per poter rendere normale l'eccezione, nel caso di persone con disabilità, sono state prese varie iniziative a livello internazionale. La World Intellectual Property Organisation, per esempio, supervisiona le attività dell'Accessible Books Consortium (ABC), che comprende organizzazioni che rappresentano disabili, biblioteche, enti normativi, scrittori, editori e organismi collettivi di diverso genere. Lo scopo primario dell'ABC è quello di espandere il numero di libri accessibili e delle opere letterarie a favore delle persone con disabilità visive¹⁷, assecondando, a livello internazionale, l'applicazione del principio del *fair use*, che legittima molti utilizzi personali o di studio, non contrastanti i diritti di editori e autori, senza particolari autorizzazioni.

2.1. Il panorama italiano

Non ci sono studi recenti specialistici di ambito LIS nel panorama italiano riguardo alle disabilità visive e all'accesso all'informazione, ma parlando in generale di terza missione universitaria e del contributo che le biblioteche possono dare alle categorie svantaggiate, è interessante l'approccio proposto da alcuni componenti della

¹⁵ Abhay Chaurasia – Ajay Pratap Singh, *Assistive support through technologies for persons with disabilities in libraries*, «DESIDOC Journal of Library & Information Technology», 42 (2022), n. 2, <<https://doi.org/10.14429/djlit.42.2.17558>>.

¹⁶ Stephen Mutula – Rebecca M. Majinge, *Information behaviour of students living with visual impairments in university libraries: a review of related literature*, «The Journal of Academic Librarianship», 42 (2016), p. 522-528, <<http://dx.doi.org/10.1016/j.acalib.2016.06.019>>.

¹⁷ Laurence Diver – Aline Iramina – Burkhard Schafer, *Exceptions for disability*, Copyrightuser.org, <<http://tinyurl.com/kxhx8sud>>.

Commissione nazionale biblioteche delle università e della ricerca dell'Associazione italiana biblioteche (AIB-CNUR)¹⁸.

Il patrimonio bibliografico e culturale di una biblioteca dell'università è legato strettamente alla storia e alle vicende dell'ateneo in cui risiede, dove si accresce e dove è fruito, ma spesso esso è correlato alla storia dello sviluppo del patrimonio di altre istituzioni del territorio in cui si trovano biblioteca e ateneo. Da questa rete di collegamenti tra enti diversi possono nascere molti progetti di valorizzazione delle risorse culturali che non escludano affatto le categorie disabili dalle diverse occasioni di crescita conoscitiva. Per esempio, tale è l'approccio del progetto *Storie di libri e palazzi*, itinerario culturale che il Sistema bibliotecario dell'Università di Ferrara ha inaugurato nel 2018¹⁹. Il progetto è rivolto ai pubblici con disabilità, e indirizzato allo sviluppo dell'inclusione – che è valso il riconoscimento del Premio AIB Maria Abenante, relativo ai progetti sull'inclusione in tutti i suoi aspetti. Il senso di tale itinerario *biblioturistico*, quindi, è stato quello di evidenziare come solo facendo rete tra biblioteche accademiche e altri istituti si possono veramente mettere a frutto pratiche e competenze. Ed è sulla base di tali reti che il contesto italiano dovrebbe muoversi nel condividere progetti, strumenti e competenze per i programmi dedicati all'inclusione e agli utenti con disabilità, non solo visive²⁰. In questo progetto, fruibile anche a distanza, l'itinerario è impostato in modo da poter coinvolgere pubblici molto diversi, con maggiore attenzione ai giovani, alle scuole e alle disabilità. Ai totem di guida degli itinerari, per esempio, sono stati apposti sensori Near Field Communication (NFC) che consentono anche ai pubblici con disabilità visiva di ascoltare audioguide attraverso alcune app. L'apertura al mondo della disabilità si era, comunque, già manifestata in altro ambito, tramite la pubblicazione di 32 video nella dat (LIS), realizzati grazie alla collaborazione del Sistema bibliotecario di ateneo ferrarese con il Coordinamento alle politiche delle pari opportunità e disabilità dell'ateneo e con l'Associazione per l'inclusione di udenti e sordi (Aidus)²¹. È stata proprio la dinamica collaborativa che ha consentito una concreta formazione reciproca condivisa fra università, scuola e terzo settore, attraverso la mediazione dei bibliotecari, al fine di creare inclusione e strategie comunicative per i diversi pubblici.

A livello nazionale, importante è ricordare l'attività del Gruppo di studio sull'inclusione dell'AIB, il quale si occupa a tutti i livelli di inclusione e di studiare e definire pratiche e linee guida per la vera "apertura" delle biblioteche a qualsiasi categoria svantaggiata di persone²².

¹⁸ Cristina Baldi [et al.], *Biblioteche accademiche in rete per la terza missione? Valorizzazione, inclusione e nuove opportunità strategiche*, «AIB studi», 62 (2022), n. 3, p. 639-652, <<https://doi.org/10.2426/aibstudi-13782>>.

¹⁹ *Storie di libri e palazzi. Itinerari fra le biblioteche di Unife*, <<http://tinyurl.com/bdfmx58v>>.

²⁰ Maria Cassella, *Biblioteche accademiche e terza missione*, Milano: Editrice bibliografica, 2020.

²¹ Università degli studi di Ferrara. Sistema bibliotecario di ateneo, *Storie di libri e palazzi: video LIS (lingua italiana dei segni)*, <<http://tinyurl.com/mwtte8mp>>.

²² Associazione Italiana Biblioteche, Gruppo di studio sull'inclusione: <<https://www.aib.it/struttura/gruppo-studio-inclusione/>>.

Altri fruttuosi esempi di dinamiche collaborative sono riconducibili a diversi programmi per le disabilità visive, sviluppati fuori dalle università, ma con esse condivisi:

Cataloghi di libri in formato accessibile

Alphabeticata – Portale bibliografico ICCU con possibilità di ricerca di risorse accessibili

<<https://alphabeticata.it/web/alphabeticata/>>

Centro Internazionale del libro parlato

<<https://www.libroparlato.org/ricerca-audiolibri/>>

Centro nazionale del libro parlato Francesco Fratta

<<https://www.libroparlatoonline.it/catalogo/cerca/1/1>>

Biblioteca Italiana per i ciechi Regina Margherita – ONLUS

<<https://www.bibliotecaciechi.it/cataloghi/cataloghi/>>

Elenco testi Università di Brescia – Sviluppato per indicare le risorse accessibili

<https://docs.google.com/spreadsheets/d/1lw6nr_BQ2NHcCS9IZ4q6B4ZAEqPQvmojJ4a8rt4e6Y/edit#gid=0>

Associazioni di sostegno

Libro parlato Lions

<<https://libroparlatolions.it/>>

Tutti i servizi per i non vedenti e per i dislessici sono gratuiti.

Biblioteca nazionale Marciana

<<https://bibliotecanazionalemarciana.cultura.gov.it/la-biblioteca/servizi/servizio-libro-parlato-lions>>

La Biblioteca Marciana offre un servizio dedicato ai non vedenti, ipovedenti e dislessici nell'ambito del progetto Libro parlato Lions.

Centro internazionale libro parlato (CILP) Adriano Sernagiotto

<<https://www.libroparlato.org/>>

<<https://www.facebook.com/LibroParlatoFeltre/>>

Pagine web ad alto contrasto. Servizio riservato a persone con disabilità visiva, con disturbi specifici di apprendimento, anziani con difficoltà di lettura autonoma,

persone con difficoltà di lettura autonoma o disagio per problematiche psico-fisiche, di devianza, di degrado o di emarginazione sociale.

Centro nazionale del libro parlato online (Unione italiana dei ciechi e ipovedenti)
<<https://www.libroparlatoonline.it/>>

Il servizio si rivolge a tutti gli utenti che abbiano una disabilità visiva. Il servizio è gratuito.

Associazione italiana dislessia (AID)
<<https://www.aiditalia.org/>>

Servizio che consente ai ragazzi con diagnosi DSA o certificazione 104/92 di richiedere il formato digitale (file PDF aperto) di testi scolastici della scuola primaria, secondaria di I e II grado, grazie al protocollo siglato con l'Associazione italiana editori (AIE). Gli editori concedono i file pdf dei testi, gratuitamente, a eccezione di quelli universitari, di narrativa e dei vocabolari.

FISH Lazio
<<https://www.fishonlus.it/lazio/>>

Organizzazione ombrello tra associazioni impegnate, a livello nazionale e locale, in politiche mirate all'inclusione sociale delle persone con differenti disabilità.

3. Le fasi del progetto

3.1. Attività preliminari

Fin dall'inizio del progetto *I libri Sapienza parlano* sono stati presi in considerazione gli aspetti catalografici delle risorse in formati accessibili e della modalità con cui renderli disponibili alle persone richiedenti. L'articolo 71-bis della legge n. 633/1941, al comma 2-undecies definisce tra gli obblighi ai quali devono adempiere le entità autorizzate anche quello di fornire, in modo accessibile, «l'elenco delle opere [ossia dei libri] o di altro materiale per cui dispongono di copie in formato accessibile e i formati disponibili». In questo modo gli utenti che ne hanno diritto possono verificare cosa sia disponibile e se lo sia nel formato per loro accessibile, mentre le altre istituzioni che godono della medesima eccezione possono chiedere lo scambio di copie, secondo i commi 2-decies e 2-undecies. In Sapienza questo obbligo di legge ha rafforzato la convinzione della necessità di descrivere nel Catalogo biblioteche Sapienza²³ i PDF e gli audiolibri realizzati piuttosto che realizzare e aggiornare liste pubblicate sulla pagina web del Sistema bibliotecario. Purtroppo, al momento, mancano indicazioni su come descrivere le risorse che presentano requisiti di accessibilità²⁴. Standard internazionali, regole nazionali e

²³ <https://opac.uniroma1.it/>.

²⁴ *Creating and editing accessibility metadata MARC tags for library staff*,
<<https://accessiblelibraries.ca/resources/accessibility-metadata-for-library-staff/>>.

guide alla catalogazione non presentano né set di dati bibliografici, né indicazioni sulle modalità più efficaci per permettere agli utenti di recuperare o identificare facilmente le risorse bibliografiche corrispondenti alle loro necessità. In particolare, mancano i campi a lunghezza fissa con vocabolari controllati che permettono di esprimere in modo uniforme e costante le caratteristiche delle risorse accessibili. Attraverso questi campi e vocabolari controllati è possibile permettere ricerche mirate o, a seguito di una ricerca generica, selezionare i risultati. Lo standard ISBD dell'IFLA presenta un elemento promettente che, in combinazione con altri, potrebbe costituire il primo passo di una ricerca finalizzata a individuare le risorse accessibili: si tratta della sensorialità dell'area 0 tramite la quale è possibile indicare il senso o i sensi coinvolti nella fruizione della risorsa bibliografica. Il formato bibliografico MARC21 presenta invece il campo 341 Accessibilità del contenuto nel quale è possibile indicare sia il senso coinvolto nella fruizione della risorsa che le caratteristiche di accessibilità presenti (ad esempio una risorsa per la cui fruizione si utilizza l'udito provvista di sottotitoli, o una risorsa visuale con audiodescrizione)²⁵. In mancanza, quindi, di indicazioni specifiche, il Centro sistema bibliotecario della Sapienza ha provveduto a ideare un modello descrittivo, valido solo a livello del Polo RMS, per i PDF accessibili²⁶ e per gli audiolibri²⁷. Queste le caratteristiche principali delle scelte descrittive:

- nell'area dell'edizione appare l'indicazione «Copia accessibile realizzata ai sensi della L. 633/1941, art. 71-bis, comma 2-ter». La formulazione di edizione originaria, se presente, è aggiunta nell'area delle note;
- nell'area della pubblicazione viene inserita la formulazione «Roma : Sapienza». La scelta è dettata dal fatto che le risorse accessibili sono realizzate da Sapienza. La formulazione originaria viene indicata nell'area delle note;
- la creazione, sempre solo in Polo, di una pseudo-collana editoriale (creata come titolo RD Raccolta di documenti) dal titolo "Libri accessibili Sapienza (ai sensi L. 633/1941, art. 71-bis)". La collana è un espediente utile per riunire insieme tutte le descrizioni e permettere di recuperarle nel Catalogo tramite un unico click;
- nell'abstract (campo usato impropriamente in mancanza di uno specifico) vengono fornite informazioni relative al progetto, ossia, «Versione realizzata nell'ambito del progetto di terza missione di Sapienza università di Roma *I libri Sapienza parlano*, ai sensi della L. 633/1941, art. 71-bis, comma 2-quinquies e seguenti», e la lista delle categorie autorizzate a richiedere le copie accessibili.

Al posto della collocazione viene inserito un cosiddetto *esemplare digitale*, costi-

²⁵ Nel 2018 sono stati introdotti due campi nel MARC21 bibliografico, il 341 *Accessibilità del contenuto* e il 532 *Nota sull'accessibilità*, e due sottocampi per lo 041 *Codice della lingua*, il \$q *Codice della lingua dell'audio accessibile* e \$r *Codice della lingua del linguaggio visivo accessibile*: <<https://www.loc.gov/marc/bibliographic/>>.

²⁶ Esempio di descrizione di un PDF: <<http://tinyurl.com/27kcar96>>.

²⁷ Esempio di descrizione di un audiolibro: <<http://tinyurl.com/mrxtr3e>>.

tuito da un link alla pagina web del progetto dove si spiega come richiedere le copie accessibili. Questa soluzione è stata adottata per sopperire al fatto che la copia digitale non può essere scaricata o fruita direttamente dagli utenti, dal momento che è necessario prima verificarne il possesso dei requisiti di legge.

Coerentemente con le politiche di cura del patrimonio digitale culturale sviluppato dalle biblioteche Sapienza²⁸, i libri digitalizzati e i relativi derivati in formato accessibile verranno immagazzinati nel sistema di conservazione della Sapienza Digital Library (SDL). La struttura di conservazione è un sistema informatico (Rosetta)²⁹, conforme allo standard ISO di conservazione Open Archival Information System (OAIS)³⁰, che mette Sapienza nelle condizioni di assolvere agli obblighi previsti per le entità autorizzate.

La gestione dell'insieme delle misure di protezione delle opere sotto diritti e dei processi abilitanti alla loro fruizione in formato accessibile sono in via di implementazione, mediante l'integrazione delle funzionalità native del sistema di conservazione con quelle del catalogo Sapienza e con quelle di altri sistemi, al momento solo individuati, per la conversione automatica e la gestione di testi e materiale didattico in formato accessibile. I libri digitalizzati e i relativi formati accessibili sono corredati di tutta la documentazione amministrativa comprovante la relazione giuridica tra il libro originale sotto diritti e la legittimazione a conservare i corrispondenti oggetti digitali. Una volta caricati nel sistema di conservazione di SDL gli oggetti digitali sono protetti da accessi non autorizzati e resi disponibili «unicamente ai beneficiari» o eventualmente «ad altre entità autorizzate», secondo l'art. 71-bis, comma 2-undecies, in base al modello di autorizzazioni del sistema, definito ad hoc per questa tipologia di oggetti.

Il sistema di conservazione SDL, preservando e proteggendo gli oggetti digitali nel lungo periodo, mantiene i dati di gestione della catena di custodia, attraverso la tenuta dell'integrità che abilita anche le garanzie di autenticità degli stessi. Inoltre, nativamente registra le operazioni effettuate sugli oggetti digitali, amministra la relazione tra gli oggetti e, pertanto, facilita sia la gestione che l'accesso dei diversi formati. La gestione conservativa dei libri accessibili, infine, riduce il rischio di operazioni ridondanti e abilita il riuso degli oggetti derivati con il conseguente miglioramento e la velocizzazione delle procedure di accesso ai formati accessibili di opere già esistenti nel sistema. Di conseguenza, nel lungo periodo, si ottiene un miglioramento complessivo del servizio agli utenti che hanno la necessità di fruire dei formati accessibili. Analogamente, lo scambio con altre entità autorizzate viene facilitato sia per la registrazione delle informazioni di scambio che per la gestione

²⁸ *Sistema bibliotecario Sapienza 2012-2020*, a cura di G. Solimine, E. Tarantino, Roma: Sapienza Università editrice, 2020, p. 119-132, <<https://doi.org/10.13133/9788893771603>>.

²⁹ Ex Libris Rosetta:

<https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation/940_Version_8.0>.

³⁰ Giovanni Michetti, *Il modello OAIS*, «*Digitalia*. Rivista del digitale nei beni culturali», 3 (2008), n. 1, p. 32-49, <<https://digitalia.cultura.gov.it/article/view/441>>.

e l'aggiornamento delle informazioni previste per le categorie di beneficiari (art. 71-bis, comma 2 duodecies).

Per la realizzazione del progetto si è scelto di attingere al ricco patrimonio e alle competenze del personale delle biblioteche Sapienza. Fin da subito, si è deciso di non adottare un approccio unilaterale, per affidarsi, al contrario, a una visione flessibile che individuasse formati differenti, nella convinzione che non esiste una forma unica per rendere accessibile i contenuti scritti, ma che la soluzione più adatta dipende da una molteplicità di fattori, quali tipo di contenuto, tipo di utilizzo, tipo di utente. Si è quindi deciso che i testi sarebbero stati resi accessibili attraverso tre principali modalità: la scansione, la realizzazione di audiolibri, la trasposizione di alcune parti in tavole tattili. Si è stabilito di valutare la modalità migliore considerando la specificità di ogni volume e il tipo di utilizzo. Ad esempio, si è deciso di scegliere la scansione per testi piuttosto lunghi, mentre si è optato per il formato audio per quei libri che si è ritenuto potessero essere fruiti per piacere più che per studio, mentre per le illustrazioni che fornivano informazioni fondamentali a integrare il contenuto testuale si è proceduto a testare la realizzazione di tavole tattili.

Nelle attività sono stati coinvolti diversi partner, tra cui l'Istituzione sistema biblioteche centri culturali di Roma Capitale³¹ e Il Centro internazionale del libro parlato³², con i quali è stato possibile diffondere la notizia del progetto, organizzare la giornata iniziale di presentazione, organizzare gli incontri formativi per i donatori di voce. Il rapporto con i partner sarà mantenuto in futuro anche al fine di mettere in atto una procedura che consenta agli utenti finali di richiedere la conversione di testi. Tra le attività preliminari del progetto sono state prese in considerazione la scelta dei formati, dei supporti e dei testi da rendere accessibili. In fase di avvio del progetto la scelta è stata effettuata dai bibliotecari, con l'intenzione di sviluppare successivamente, una volta che le procedure fossero entrate a regime, un processo di selezione condivisa attraverso lo scambio circolare di esperienze, conoscenze e richieste tra bibliotecari, docenti e utenti, con la prospettiva che gli utenti, attraverso le associazioni di riferimento, possano esprimere i propri desideri.

Formati e supporti

3.2. Audiolibri

La lettura ad alta voce è un'attività dalle molteplici sfaccettature e con diverse possibilità di impiego³³. I cambiamenti culturali e tecnologici, che hanno attraversato nel corso del tempo il mondo della lettura, non hanno cancellato bensì moltiplicato i significati messi in comune dall'incontro voce umana/ascoltatore. Nei

³¹ <https://www.bibliotecheidiroma.it/opac/.do>.

³² <https://www.libroparlato.org/>.

³³ Un cenno al progetto *I libri Sapienza parlano* è presente nel contributo Valentina Rovacchi – Elisabetta Tamburini – Antonella Fallerini, *Biblioteche accademiche ad alta voce*, in: *La lettura ad alta voce condivisa. Atti del primo convegno scientifico internazionale, Perugia 1-2 dicembre 2022*, a cura di F. Batini, G. Marchetta, Lecce: Pensa multimedia, 2023, p. 287-288.

contesti più disparati, educativi, riabilitativi, assistenziali, una voce messa al servizio di chi ascolta (con le sue caratteristiche sonore, di pronuncia, di tono), contribuisce in un'ottica di inclusione a stabilire una relazione forte, capace di ridurre lo svantaggio iniziale.

L'audiolibro ha seguito percorsi di sviluppo diversi, di natura commerciale, con l'aumento di una produzione che, appunto perché pensata per un qualsiasi tipo di lettore-ascoltatore, è stata sempre più protetta da diritti di utilizzo che non tengono conto delle esigenze, per esempio, di una biblioteca che li raccoglie per farli ascoltare ai propri utenti con disabilità visive. Problema che non si è risolto nemmeno con i libri prodotti appositamente da gruppi di assistenza, sempre tenuti a seguire le restrizioni del copyright³⁴. In entrambi i casi, comunque, per quanto riguarda la specifica utenza che dovrà ascoltare i libri, sono state poco investigate le preferenze riguardo al genere, o al tono, o all'accento di chi produce l'audiolibro, tutte variabili che si riscontrano in contesti sociali, o di età, o di luogo differenti³⁵.

Per la finalità del progetto *I libri Sapienza parlano* si è stabilito di selezionare alcuni libri da affidare a volontari *donatori di voce*, ritenendo che l'ascolto della lettura da parte di una voce umana fosse un'esperienza più coinvolgente, comprensibile e gradevole. A tal fine, sono stati scelti testi che avessero le caratteristiche di ampia fruibilità e non fossero tecnici o specialistici; quindi, sono stati selezionati testi divulgativi, sia in ambito scientifico che umanistico, testi trasversali a più discipline, testi teorici che potessero incontrare un interesse diffuso³⁶.

Dal punto di vista pratico, si è proceduto alla raccolta di adesioni per donatori volontari aperta alla comunità Sapienza e al territorio³⁷ ed è stato organizzato un incontro informativo per gli aderenti, in collaborazione con i partner del progetto³⁸. Nell'ambito della giornata sono state illustrate la lettura ad alta voce e le modalità tecniche per registrare. In particolare, i punti affrontati hanno riguardato una breve storia della lettura ad alta voce, le applicazioni, le tecniche e gli elementi della lettura ad alta voce: postura, respirazione, tono, volume, ritmo, dizione, punteggiatura, parola chiave, contesto, target, tempi. Durante il workshop i partecipanti hanno svolto un lavoro di gruppo, cimentandosi per la prima volta con la lettura

³⁴ Johan W. Roos, *Libraries for the blind as accessible content publishers: copyright and related issues*, «Library Trends», 55 (2007), n. 4, p. 879-916.

³⁵ Anna Hampson Lundh, *Talking books and reading children: children describing their use of talking books*. Johanneshov: Swedish Agency for Accessible Media (MTM), 2013, <<http://tinyurl.com/4myhef8e>>; Manorama Tripathi – Archana Shukla, *Use of assistive technologies in academic libraries: a survey*, «Assistive Technology», 26 (2014), n. 2, p. 105-118, <<https://www.tandfonline.com/doi/full/10.1080/10400435.2013.853329>>.

³⁶ <http://tinyurl.com/3zafamw3>.

³⁷ La call, come detto, è stata lanciata in occasione della *Giornata internazionale della disabilità 2022* sul sito del Sistema bibliotecario Sapienza, dove per un mese circa gli interessati hanno potuto compilare il modulo di adesione. Le risposte sono state 189.

³⁸ Giovanna Micaglio, responsabile del Servizio Scuola e rete Bibliopoint del Comune di Roma e Dario Sandro Vaggi, coordinatore locale per il Lazio del Centro internazionale del Libro parlato.

ad alta voce e applicando quanto illustrato precedentemente. Nella parte finale sono state date indicazioni tecniche per una buona registrazione ed è stato presentato Audacity³⁹, il sistema scelto per la registrazione degli audiolibri.

Nel complesso sono stati assegnati 46 libri, dati in prestito per tre mesi. Alla data odierna sono stati realizzati 32 audiolibri.

3.3. Scansioni accessibili

Un'altra forma di accessibilità proposta è rappresentata dalla scansione in formato PDF/A con riconoscimento OCR. Rispetto ai testi letti ad alta voce, questa forma consente una maggiore versatilità nell'esplorazione del contenuto, poiché l'utilizzo della modalità automatica di lettura ad alta voce, ormai molto diffusa e inclusa sia nei software di scrittura/lettura che nei principali browser web, consente in maniera piuttosto facile di calibrare e modulare tempi e modi di ascolto, ad esempio interrompendo e riprendendo la lettura a piacimento⁴⁰.

Un prototipo piuttosto efficace è stato il sistema DAISY (Digital Accessible Information System), creato dal DAISY Consortium, portale per lo sviluppo di strumenti, standard e best practices di pubblicazione e lettura per garantire l'accessibilità alle persone con disabilità visive, che fin dal 1994 ha permesso la lettura automatica di file di testo da parte di software in grado di interpretare parole e punteggiatura e riprodurle⁴¹. Il sistema ha avuto presto il ruolo di maggiore efficacia rispetto alle altre tecnologie assistive, differenziandosi subito dal ruolo dell'audiolibro. Tra le due tecnologie, infatti, c'è una similitudine solo apparente, dato che il primo può essere prodotto appositamente per i disabili senza ledere le leggi sul copyright, l'audiolibro invece essendo prodotto per qualsiasi tipo di pubblico è sottoposto alla legislazione sull'uso e la diffusione delle copie legalmente acquistate. Tutto questo, però, ha cominciato a perdere senso grazie ai progressi delle tecnologie per la produzione di libri digitali, che hanno portato presto alla produzione di semplici PDF leggibili automaticamente dallo stesso software gratuito che li visualizza e da molti altri⁴².

Nell'ambito del progetto *I libri Sapienza parlano*, le scansioni accessibili sono state perlopiù affidate ai volontari del Servizio civile universale dei progetti Flux⁴³ e #nonsololibri⁴⁴, che hanno collaborato con le biblioteche di Sapienza nei progetti 2023-2024, e a studenti collaboratori per la terza missione assegnati allo scopo.

³⁹ <https://www.audacityteam.org/>.

⁴⁰ Anna Hampson Lundh – Genevieve Marie Johnson, *The use of digital talking books by people with print disabilities: a literature review*, «Library Hi Tech», 33 (2015), n. 1, p. 54-64, <<https://doi.org/10.1108/LHT-07-2014-0074>>.

⁴¹ The DAISY Consortium: <<https://daisy.org/>>.

⁴² Ken Petri, *Accessibility issues in e-books and e-book readers*, in: *No shelf required 2: use and management of electronic books*, ed. by S. Polanka, Chicago: American Library Association, 2012, p. 35-60.

⁴³ <http://tinyurl.com/4xfmvs7s>.

⁴⁴ <http://tinyurl.com/yc754xy9>.

La scelta dei libri da scansionare è stata effettuata in fase di avvio del progetto dai bibliotecari coinvolti, che hanno selezionato principalmente testi di media lunghezza (di norma superiore alle 200 pagine), diversi dei quali indicati nei programmi d'esame dei vari corsi di laurea. Affinché le scansioni rendessero al meglio il volume cartaceo, in alcuni casi si è deciso di integrare il testo inserendo le descrizioni di immagini, che in questo modo sono andate ad arricchire notevolmente la possibilità di veicolare il contenuto informativo nella sua completezza. In questi casi, una volontaria del Servizio civile universale si è occupata di realizzare una descrizione scritta delle immagini contenute nei testi, che sono state inglobate nel file PDF ottenuto dalla scansione.

Nel corso di questa attività ci si è resi conto che in molti casi si sono dovuti escludere dalla selezione volumi in cattive condizioni di conservazione, soprattutto perché annotati e sottolineati; questo stato, infatti, non consente alla gran parte dei software di lettura automatica di riconoscere adeguatamente il testo. Tale consapevolezza, maturata sul campo per esperienza pratica diretta, ha portato all'adesione convinta alla campagna internazionale *Se tu sottolinei io non leggo*. La campagna è nata in Spagna alla fine del 2010 dalla collaborazione tra l'Oficina para la inclusión de personas con diversidad e le biblioteche dell'Università Complutense, e ha lo scopo di realizzare diverse azioni di sensibilizzazione rivolte a tutti gli utenti delle biblioteche per evitare di sottolineare i libri nell'ottica dell'accessibilità⁴⁵. Nell'arco di dodici anni l'iniziativa è stata accolta in diverse nazioni, tra cui Italia⁴⁶, Portogallo e Colombia, e nel 2022 è stata ripresa dall'Università dell'Estremadura⁴⁷. Nell'ambito del progetto *I libri Sapienza parlano* l'iniziativa è stata ripresa in collaborazione con l'Istituzione sistema biblioteche centri culturali di Roma capitale e lanciata in occasione della Giornata internazionale delle persone con disabilità del 3 dicembre 2023⁴⁸. Per l'occasione sono stati realizzati materiali grafici (poster, totem, segnalibri), che sono stati distribuiti nelle biblioteche, e bollini adesivi apposti sulle copertine dei libri. La campagna è stata diffusa sui siti e sui social delle biblioteche aderenti.

3.4 Libri tattili

Una delle maggiori difficoltà per un disabile visivo è poter apprendere e godere di quello che, con estrema generalizzazione, definiamo mondo dell'arte: per lungo tempo c'è stata la convinzione, fortemente radicata tra i vedenti e poi condivisa

⁴⁵ <http://tinyurl.com/3c6bwwfd>.

⁴⁶ Per esempio, nel 2019 dalle biblioteche dell'Università di Genova: <<https://biblioteche.unige.it/node/717>>. Si veda in proposito: Marcella Rognoni Nuccia Giannelli, *Caso di studio Se tu sottolinei io non leggo*, in: *La biblioteca accademica*, a cura di D. Deana, Milano: Editrice Bibliografica, 2022.

⁴⁷ <http://tinyurl.com/ywryvvyr>.

⁴⁸ <https://www.uniroma1.it/it/notizia/se-tu-sottolinei-io-non-leggo>.

da molti anche nelle comunità dei non vedenti, secondo la quale le immagini, considerate come il luogo del visivo per eccellenza, sarebbero in gran parte precluse a chi non vede. È innegabile che la vista consenta un processo di conoscenza della forma fluido e immediato: l'osservazione visiva dell'oggetto permette molto rapidamente, appunto con un colpo d'occhio, di indagare, misurare, collocare gli oggetti e i luoghi e offre al sistema percettivo indicazioni per organizzare osservazioni, confronti, valutazioni soggettive. Ma la vista non è l'unica e più completa possibilità di conoscenza. Lo stimolo tattile è il secondo senso più utilizzato dopo la vista, e questo può rivelarsi fondamentale per conoscere e apprezzare il valore estetico di una scultura o di una architettura conosciuta attraverso riproduzioni a rilievo: attraverso il tatto possono essere resi accessibili singoli aspetti della realtà circostante altrimenti impossibili da raggiungere se non attraverso la mediazione spesso troppo vaga e ingannevole della sola descrizione verbale.

L'esplorazione con le mani si estende, a differenza della vista, su un campo percettivo ridotto e raggiungibile, a portata di mano: procede con sequenzialità su frammenti spaziali, ma è in grado di definire, con una attenta valutazione analitica, la sensazione termica del contatto con l'oggetto e, attraverso micro movimenti delle dita, la qualità della superficie e la sua consistenza. Ma per arrivare alla comprensione dell'oggetto rappresentato, la formulazione complessiva del lavoro di illustrazione tattile deve essere frutto di un'attenta correlazione tra descrizione verbale e qualità della riproduzione: il testo deve descrivere i tratti essenziali, guidare eventualmente l'esplorazione delle mani, aggiungere altre informazioni che aiutino il lettore a cogliere quanto con le sole mani non potrà mai percepire e, insieme, interpretarne il senso; una comunicazione plurisensoriale inclusiva a tutti gli effetti.

In particolare, l'architettura, pur continuando a essere considerata un'arte visiva, coinvolge il maggior numero di sensi: un testo che abbia delle immagini da studiare pone l'esigenza, per il disabile visivo, di una riproduzione che gli consenta di vedere, ma anche valutare nello spazio, la facciata, il campanile, la cupola e gli ornamenti, o addirittura una prospettiva: partendo da queste considerazioni abbiamo deciso, all'interno del progetto *I libri Sapienza parlano*, di provare a realizzare alcuni libri tattili, prevalentemente di architettura, ma sperimentando anche con immagini di geologia e biologia, che, nelle nostre intenzioni, vorremmo far diventare una sorta di prototipo, di apripista per il futuro.

4. Design for phygital world/libri tattili

La sperimentazione si è basata su alcuni principi teorici e metodologici che fanno capo alle dimensioni *Human-centred* e *Technology-based* del design: la prima, coinvolge gli aspetti di ergonomia fisica e cognitiva, così come quelli culturali ed estetici, che fanno riferimento alla morfologia sia materiale sia immateriale, mentre la seconda presuppone la scelta e lo sviluppo di un uso opportuno dei processi tanto di progettazione che di produzione.

Il design di fatto si è sempre occupato della relazione tra le persone e l'ambiente che li circonda progettando le cose materiali o immateriali che si offrono come medium: in altre parole il design si è sempre occupato, attraverso la forma, di esperienza. E nel linguaggio filosofico, l'esperienza è riconosciuta proprio come quel tipo di conoscenza acquisita tramite i sensi come contrario del puro ragionamento teorico. Da qui la particolare attenzione che il design ha dato via via ai sensi attraverso un passaggio evolutivo che, partendo dal concetto di godimento/appagamento estetico, proprio della dimensione di arte applicata del proto-design⁴⁹, ha attraversato la dimensione scientifica della Gestalt⁵⁰, quella deterministica dell'ergonomia⁵¹, per poi approdare ai concetti inclusivi e *Human-centred* dell'emozione/sinestesia⁵². Il design approda alla sinestesia quando prende in considerazione proprio quel fenomeno psicofisico della possibilità di avere una percezione di natura sensoriale diversa da quella relativa al senso effettivamente sollecitato, come per esempio la percezione di un'immagine visiva in seguito a uno stimolo acustico o tattile o olfattivo. In questo modo non si limita alla natura biologica e fisica della persona, ma riflette sulla sua natura cognitiva, emozionale e quindi immateriale. Questa evoluzione ha comportato un graduale passaggio dal *progetto per la disabilità*, dove l'attenzione era quella al superamento di tutti gli impedimenti di carattere funzionale specificatamente pensato per i bisogni di una categoria di persone, al *progetto per l'inclusività* dove ogni utente può trovare risposta alle proprie esigenze siano esse riconducibili alla sfera della *normalità* come condizione più comune che a quella della *diversa normalità* come condizione più specifica. Questa particolare sensibilità ha dato vita a una filosofia progettuale per la diversità umana, l'inclusione sociale e l'uguaglianza⁵³ denominata *Design for All*, il cui scopo è facilitare le «pari opportunità di partecipazione in ogni aspetto della società» attraverso una corretta progettazione degli ambienti, degli oggetti, dei servizi, così come della cultura e delle informazioni, superando, attraverso l'inclusione, le diverse abilità fisiche o cognitive.

Da questi presupposti metodologici ha preso le mosse la sperimentazione progettuale di *Design for phygital world/libri tattili*, con il tentativo di cimentarsi con un approccio *Design for All* al fine di trovare una sintesi sinestetica tra la percezione visiva e quella tattile. Come è noto tanto la percezione visiva che quella tattile sono temi di ricerca e sperimentazione consolidati che hanno visto l'impegno di figure di

⁴⁹ Anty Pansera, *Cose belle e utili: la nascita del prodotto industriale*, in: *Enciclopedia Treccani*, 2013, <[⁵⁰ Michele Sinico, *Gestaltismi e fenomenologie nella Bauhaus*, in: *Ragionamenti percettivi. Saggi in onore di Alberto Argenton*, a cura di C. M. Fossaluzza, I. Versteegen, Milano: Mimesis, 2014.](https://www.treccani.it/enciclopedia/cose-belle-e-utili-la-nascita-del-prodotto-industriale_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Tecnica)/>.</p></div><div data-bbox=)

⁵¹ Francesca Tosi, *Ergonomia e progetto*, Milano: Franco Angeli, 2013.

⁵² Donald A. Norman, *Emotional design*, Milano: Apogeo, 2004.

⁵³ Avril Accolla, *Design for All*, Milano: Franco Angeli, 2009.

spicco non solo nel campo della psicologia o della medicina, ma anche nel campo dell'arte, come Arnheim, della pedagogia, come Montessori, del design, come Munari. Ovviamente le due categorie percettive sono state principalmente studiate come campi di indagine separate con lo scopo di comprenderne il ruolo nello sviluppo e nelle dinamiche cognitive delle persone. E in qualche modo, mentre la percezione visiva è stata trattata come medium primario con cui la persona costruisce il proprio bagaglio di conoscenza/esperienza, per quella tattile si è cercato maggiormente di indagarne gli aspetti sinestetici. Basti pensare, in campo pedagogico, ai libri tattili sia di Montessori che di Munari il cui scopo è quello di introdurre e allenare il bambino alla comprensione degli stimoli esterni e alla loro collocazione cognitiva. In una ulteriore direzione sono invece andati gli studi legati alla percezione tattile di persone con disabilità visive, correttamente concentrati nello sviluppo e nella sperimentazione di strumenti e nuovi linguaggi di comunicazione che sopperissero alla mancanza del senso della vista. Emblematica è la storia dell'alfabeto Braille, messo a punto dal francese Louis Braille nella prima metà del XIX secolo e ancora oggi strumento principale di trasmissione di informazioni per i non vedenti e gli ipovedenti.

Sulla base di queste premesse è nata la sperimentazione *Design for phygital world/libri tattili*, in collaborazione con il Laboratorio di design per la fabbricazione digitale Sapienza Design Factory⁵⁴, con lo scopo di indagare progettualmente:

- il bisogno di conoscenza espresso dalle persone con una disabilità visiva (totale e parziale) che, per evidenti ragioni, non possono attingere al patrimonio bibliografico che fino ad oggi è stato pensato esclusivamente per il principale dei nostri sensi, la vista;

- le opportunità offerte dalla fabbricazione digitale per sviluppare artefatti percettibili attraverso un altro dei nostri sensi, il tatto, grazie anche alla possibilità di personalizzazione delle morfologie.

La sperimentazione ha visto il coinvolgimento di sei studentesse, di uno studente e tre docenti⁵⁵ del corso di laurea triennale in Design della Sapienza. Il lavoro ha

⁵⁴ Sapienza Design Factory è un laboratorio di ricerca sperimentale del Dipartimento di Pianificazione, design, tecnologia dell'Architettura della Sapienza. È responsabile tecnico il dott. Marco Cavallini Chialastri, responsabile scientifico è la prof.ssa Loredana Di Lucchio.

⁵⁵ Hanno collaborato al progetto in qualità di studenti/designer: Giacomo Corvina, Fabiana Gentili, Alessandra La Penna, Alina Mihai, Alessia Miglietta, Ludovica Orlandi. I docenti tutor del progetto sono stati i prof. Andrea Casale (Disegno), Loredana Di Lucchio (Design e Responsabile del progetto), Francesco Di Nocera (Psicologia). Nello specifico la studentessa Fabiana Gentili ha lavorato sul testo scelto dalla Biblioteca del Dipartimento di Scienze della Terra (Direttore Daniele Coltellacci) *Introduzione alla cristallografia*, di Germano Rigault; le studentesse Alessandra La Penna e Alessia Miglietta hanno lavorato sul testo scelto dalla Biblioteca centrale della Facoltà di Architettura (Direttrice Laura Armiero) insieme alla Biblioteca del Dipartimento di storia, disegno e restauro dell'Architettura (Direttrice Alessandra Gulotta) *Roma Barocca*, di Paolo Portoghesi; le studentesse Alina Mihai e Ludovica Orlandi e lo studente Giacomo Corvina hanno lavorato sul testo scelto dalla Biblioteca "Ludovico Quaroni" (Direttrice Sveva Barbera) del Dipartimento di Pianificazione, design, tecnologia dell'Architettura, *Comunicazione Visiva*, di Bruno Munari.

avuto come oggetto tre libri opportunamente scelti dalle biblioteche coinvolte e, attraverso uno studio delle problematiche legate alla lettura visiva e a quella tattile, è stata immaginata e realizzata la versione tattile di alcune figure, presenti a corredo iconografico dei testi indagati. I prototipi, realizzati sfruttando una tecnologia di fabbricazione digitale (additiva e/o sottrattiva), sono stati testati con utenti con disabilità visiva per vanificarne la funzionalità e recepire informazioni utili al miglioramento di ognuna delle tavole tattili. Il lavoro di sperimentazione è stato quindi organizzato in tre fasi tra loro consequenziali: la fase uno dedicata all'approfondimento e verifica delle competenze necessarie ad affrontare il percorso di progettazione, con riferimenti ai temi da una parte della percezione visiva e dall'altra della fabbricazione digitale; la fase due, orientata alla conoscenza dell'apparato iconografico del testo oggetto di studio e della sua interpretazione attraverso codici tattili adeguati; la fase tre, volta allo sviluppo del prototipo e al test con gli utenti reali.

La sperimentazione, svolta con l'obiettivo di permettere la comprensione di contenuti visivi senza l'uso del senso della vista, ha dunque tenuto conto e approfondito entrambi i fenomeni percettivi per trovare dei punti di divergenza e stabilire regole di convergenza. Inoltre, i libri su cui è stata fatta la sperimentazione, essendo testi universitari, presentano una particolare categoria di corredo iconografico, strettamente immaginato per la comprensione dei contenuti testuali, e quindi prettamente didascalico, che non poteva divergere, nella sua trasposizione tattile, verso una dimensione allegorica. La progettazione è quindi partita dalla comprensione del testo da un punto di vista scientifico, conoscendone e comprendendone i contenuti, anche visuali. A partire dalle categorie cognitive astratte che gli stimoli visivi delle immagini determinavano, si è costruito un vocabolario di altrettanti stimoli tattili, corrispondenti alle stesse categorie cognitive: a solo titolo esemplificativo, alcune delle categorie cognitive indagate dal punto di vista della percezione visiva e tattile sono state la distanza (lontano/vicino), la dimensione (alto/basso, largo/stretto), il movimento e la velocità, l'assenza e la presenza. Questo ha permesso definire una sorta di codice tattile per la trasposizione non semplicemente delle immagini, ma dei concetti espressi dai testi e dalle immagini insieme. Si è poi passati alla selezione di una o più tavole iconografiche tra quelle presenti in ognuno dei libri, e si è progettata la loro trasposizione tattile usando i codici del nuovo vocabolario percettivo. Infine, sono stati sviluppati i prototipi.

4.1. 3D printing tra dimensione immateriale e materiale

Per la realizzazione dei prototipi, grazie al supporto del Laboratorio Sapienza Design Factory, è stata scelta come tecnologia di realizzazione quella della stampa 3D proprio per la semplicità e velocità di trasposizione dal progetto all'oggetto che questa offre. Un'ulteriore motivazione a supporto della scelta di questa tecnologia di fabbricazione digitale è stata la consapevolezza che essa rappresenta la

soluzione più praticata in tutte le esperienze attualmente in atto sul tema della percezione tattile di contenuti visivi, soprattutto artistico-museali⁵⁶. La stampa 3D, infatti, grazie al suo potenziale di parametrizzazione, offre la possibilità di partire da una ricostruzione virtuale dell'immagine di partenza e realizzarne una copia fisica senza vincoli dimensionali rispetto alle esigenze tattili. In particolare, è possibile riconoscere, in questa esperienza di utilizzo della stampa 3D a favore della percezione di contenuti visuali da parte di persone non vedenti (o ipovedenti), quattro categorie di azione/risultato:

- l'azione descrittiva, quando la stampa 3D permette di generare oggetti tattili da puri oggetti visuali (immagini) di cui l'utente non ha una precedente esperienza;
- l'azione di riproduzione, quando invece si tratta di replicare qualcosa che ha già una dimensione tattile, ma non è accessibile per dimensione o collocazione;
- l'azione di chiarimento, quando il contenuto è solamente astratto (come un testo, un suono, un colore) e quindi non ha una dimensione fisica di riferimento;
- l'azione interpretativa, quando seppur esistano riferimenti morfologici per il contenuto, si opta per un diverso codice morfologico per aumentarne l'intelligibilità.

Nei progetti sviluppati, queste categorie sono state tutte messe in campo in maniera sinergica per rispondere alle esigenze che i diversi contenuti scientifici dei libri ponevano al momento della lettura. L'utilizzo della stampa 3D ha permesso di realizzare una serie di prototipi via via evolutivi, tutti testati da soggetti non vedenti ma anche da persone vedenti per comprenderne l'efficacia e acquisire considerazioni e suggerimenti in un processo di *co-design*.

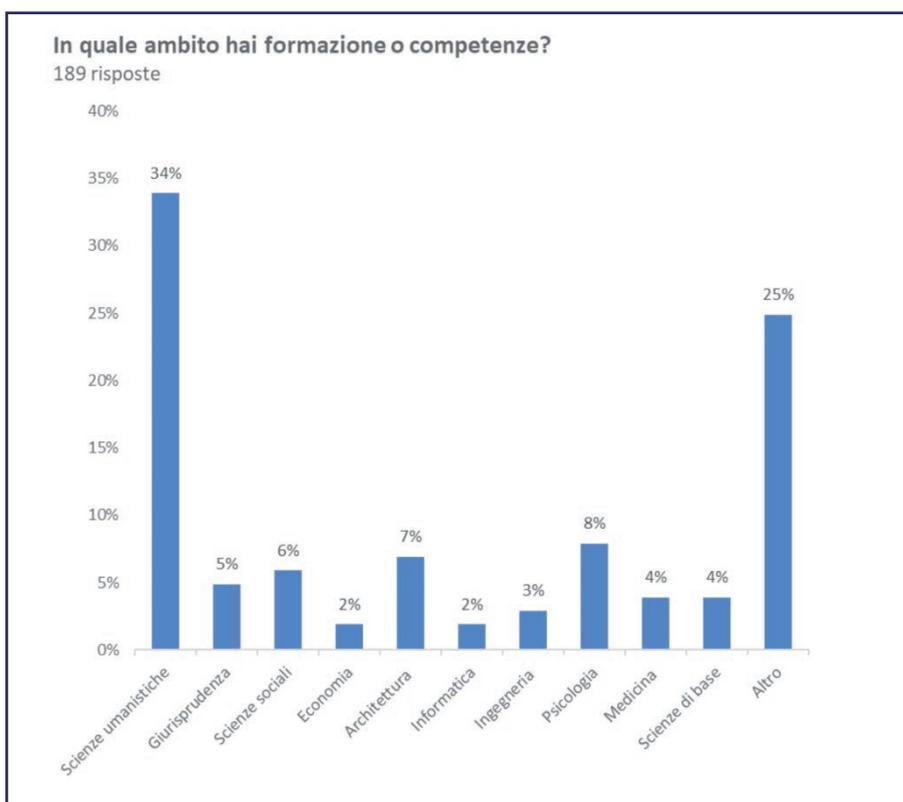
4.2. Libri tattili: dalla percezione all'espressione

Gli esiti di questa sperimentazione sono in via di conclusione e saranno messi a disposizione del Sistema bibliotecario Sapienza. Ma certamente il progetto non può considerarsi concluso, sia per le potenzialità di trasferimento a tutto il patrimonio bibliografico dell'Ateneo, soprattutto in quei campi del sapere dove la componente visuale è determinante, ma anche perché, rispondendo all'approccio del *Design for All*, si può e si deve alimentare della reale pratica d'uso. I libri tattili rappresentano, infatti, un campo di azione del *Design for All* dove la centralità delle persone è determinante: essi rappresentano non solo degli strumenti di apprendimento, ma anche potenzialmente strumenti espressivi di una percezione sinestetica, che fa convergere la dimensione visiva in quella tattile, in una sorta di interscambiabilità cognitiva.

⁵⁶ Fra tutti i progetti sviluppati per la fruibilità dei contenuti storico-artistici da parte di persone con disabilità visiva che hanno sfruttato la stampa 3D, si ricorda il lavoro del Museo tattile statale Omero ad Ancona.

5. Dati sui donatori di voce

Di seguito sono riportati i grafici realizzati utilizzando i dati forniti dagli aspiranti volontari donatori di voce al momento dell'adesione al progetto.



6. Conclusioni. Per una editoria inclusiva

Siamo vicini ormai alla data del 28 giugno 2025, a partire dalla quale entra in vigore la direttiva dell'Unione europea denominata *The European Accessibility Act* che stabilisce obiettivi vincolanti di accessibilità per molte tipologie di prodotti e servizi, tra cui quelli editoriali. La direttiva dà seguito e completa quanto già convenuto nel 2013 nel cosiddetto *Trattato di Marrakech* per facilitare, in deroga alle norme sul diritto d'autore, l'accesso alle opere pubblicate e destinate alle persone non vedenti, con disabilità visive o con altre difficoltà. Tant'è che alcune università (in verità, assai poche), tra cui la Sapienza, per il tramite dei loro Sistemi bibliotecari e del Ministero della Cultura, stanno sottoscrivendo con l'Associazione Italiana Editori i relativi accordi attuativi. Il contesto legislativo internazionale spinge, dunque, verso la creazione di un sistema editoriale accessibile, confortato in questo dalla disponibilità di standard di accessibilità per la creazione, la fruizione, la distribuzione e la vendita di opere in formato digitale; standard che hanno richiesto appunto una specifica regolamentazione per la loro applicazione. Ma quel che si impone adesso è un cambio di paradigma che investe prioritariamente il sistema universitario e, nello specifico, le case editrici delle università.

Investe il sistema universitario perché per la piena attuazione dello *European Accessibility Act* è necessaria un'assunzione di responsabilità e l'elaborazione di un piano strategico nazionale che comporta un insieme significativo di interventi strutturali, finanziari e organizzativi, nonché la promozione di una campagna di sensibilizzazione culturale sul tema dell'accessibilità ai prodotti della ricerca che, come è già avvenuto con l'open access, coinvolga tutti i ricercatori. Si richiederà perciò un significativo investimento finanziario, sul piano innanzitutto della formazione, direi meglio dell'educazione alla cultura editoriale dell'inclusione e dell'accessibilità. Ma gli investimenti dovrebbero essere canalizzati anche per stimolare e finanziare la ricerca applicata nell'elaborazione e nella creazione di formati, interfacce, piattaforme editoriali che rispondano ai bisogni e alle esigenze di studenti e studiosi disabili o con DSA. Esempari in tal senso sono le sperimentazioni condotte presso il Laboratorio Polin dell'Università degli studi di Torino e il progetto di terza missione *I libri Sapienza parlano* promosso dal Sistema bibliotecario della Sapienza Università di Roma. Quest'ultima esperienza insegna che l'accessibilità è un percorso che investe gli attori di una complessa filiera industriale e culturale, il cui successo dipende dal coinvolgimento di numerose competenze e dall'interazione tra istituzioni pubbliche e imprese. Se per alcuni prodotti editoriali le soluzioni tecnologiche sembrerebbero essere a portata di mano (audiolibri, lettura automatica di file di testo da parte di software in grado di interpretare parole e punteggiatura e di riprodurle), per altri (libri tattili, libri dalla struttura più complessa con tabelle, grafici, immagini, box illustrativi, glosse, ecc.) la ricerca di soluzioni integrate richiede un significativo impegno scientifico; magari pensando, perché no, a una nuova macchina-libro per un nuovo concetto di testo. Non si tratta tanto di adeguare lo status

quo alle esigenze espresse dalla cultura editoriale inclusiva, quanto forse di progettare servizi editoriali innovativi che sappiano rispondere dinamicamente alle esigenze della ricerca e della didattica. Un cambio di passo – dunque – che investe direttamente le case editrici universitarie perché, essendo esse uno dei principali canali di comunicazione della produzione scientifica delle università, dovrebbero farsi carico di riprogettare e riorganizzare l'intera filiera produttiva, coinvolgendo in questo processo trasformativo innanzitutto gli autori. L'integrazione tra dispositivi tecnologici per la lettura, piattaforme didattiche e piattaforme di distribuzione per la fruizione di contenuti accessibili richiede, inoltre, la disponibilità da parte del mondo dell'editoria, che deve interpretare come un'opportunità quella che viene considerata ancora, per certi versi, una minaccia.

In questa triangolazione di azioni, un ruolo centrale lo giocano le biblioteche, come dimostra la virtuosa esperienza messa in atto dal Sistema bibliotecario della Sapienza. La centralità del ruolo è data non solo e non tanto dal farsi mediatori nelle transazioni con gli utenti, ma anche, e direi soprattutto, nell'elaborazione di strategie catalografiche volte alla descrizione formalizzata e accessibile delle risorse rivolte agli utenti con disabilità, rendendo visibili i relativi metadati ai motori di ricerca; e proponendo l'implementazione di soluzioni di DRM che non ne precludano l'accessibilità.

The paper illustrates the Third Mission project of the Sapienza Università di Roma I libri Sapienza parlano (Sapienza books speak), indicating its genesis, aims, development phases and activities carried out. The project is aimed to provide people with visual disabilities and DSA, even those not belonging to the Sapienza community, with versions of the books available in accessible formats at Sapienza libraries. Sapienza University has already obtained the authorization to make accessible copies thanks to the exception to copyright in favor of certain categories of beneficiaries.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Quel che resta di un codice: nuovi orizzonti per la catalogazione dei frammenti in MOL 2.0*

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00096

Elisabetta Caldelli

Università degli studi della Campania Luigi Vanvitelli

Lucia Negrini

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU)

La reingegnerizzazione, cui Manus Online è stato sottoposto negli ultimi due anni e la cui nuova versione MOL 2.0 è stata pubblicata a partire da dicembre 2021, ha permesso di introdurre alcune importanti novità, non ultima la possibilità di catalogare e di indicizzare i frammenti di manoscritti, di cui biblioteche e archivi italiani, pubblici e privati, sono ricchi depositi. Verranno illustrati i presupposti teorici e gli aspetti tecnici del back-end relativi alla descrizione dei frammenti in MOL 2.0.

1. Premessa

I primi contatti fra l'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU) e l'Università degli Studi di Urbino Carlo Bo sono iniziati nel luglio del 2020, quando la professoressa Alessandra Molinari scrisse una e-mail all'Area manoscritti proponendo lo studio di un metodo di interoperabilità tra il database di TextusInvisibilis e Manus Online (MOL), in modo che, nel momento in cui i dati nella scheda del progetto saranno inseriti nel suo database, i dati catalografici in esso raccolti possano essere riversati in uno step successivo anche nella scheda di Manus Online.

A questo primo contatto ne sono seguiti altri; in particolare mi riferisco a due riunioni, la prima nel luglio, la seconda nel novembre del 2022, anche con la presenza di altri interlocutori: di questi incontri tratterò diffusamente Elisabetta Caldelli nella seconda parte di questo contributo. Anticipo solo che uno dei risultati di

* Questo contributo è stato presentato in occasione del Convegno internazionale "Frammenti e frammentarietà nelle discipline umanistiche e nei beni culturali", tenutosi ad Urbino nei giorni 8-10 maggio 2023. Desideriamo ringraziare la prof.ssa Alessandra Molinari per l'invito e la dott.ssa Simonetta Buttò per l'appoggio fattivo che ha sempre dato alle evoluzioni di Manus Online. Il testo si suddivide in due parti: la presentazione del progetto e della piattaforma Manus Online in generale, dovuta a Lucia Negrini (Paragrafi 1-5) e l'illustrazione delle applicazioni implementate in questa nuova versione di MOL che riguardano i frammenti, dovuta a Elisabetta Caldelli (Paragrafo 6).

questo fruttuoso dialogo è stata la decisione di elaborare una mappatura fra la scheda MOL e quella di *Textus invisibilis*¹, al fine di studiare vie di collaborazione fra i due progetti.

Si pensa anche alla creazione di un Gruppo di Lavoro che approfondisca il dialogo fra ICCU, archivi e università su materiali presenti sia negli archivi che nelle biblioteche, in modo da creare convergenze che possano portare in un futuro anche a progetti comuni.

2. Un po' di storia...

Quello che vorrei proporvi con questo mio contributo è anche, in parte, la storia di un'idea, la realizzazione di un progetto che negli anni in cui fu concepito appariva forse a molti troppo ambizioso e velleitario, quindi difficilmente realizzabile.

Tale progetto è nato da un'idea di Angela Vinay - direttrice dell'Istituto centrale per il catalogo unico dal 1976 al 1987 -, anche su sollecitazione di studiosi e bibliotecari che auspicavano una ripresa dell'attività di catalogazione dei manoscritti conservati nelle biblioteche italiane, oramai stagnante da troppo tempo. In forza della legge che lo ha istituito l'ICCU appariva come l'organismo più idoneo a promuovere e coordinare l'attività di catalogazione dei manoscritti e a curare l'elaborazione di normative nazionali di catalogazione.

Nel giugno del 1980 ebbe luogo a Roma uno storico seminario: "Il manoscritto: situazione catalografica e proposta di una organizzazione della documentazione e delle informazioni"; da questo incontro ci si attendeva la definizione di principi sui quali predisporre una comune normativa catalografica, la formulazione di elementi per una scheda unificata da proporre a livello nazionale per la descrizione del manoscritto.

La costituzione di una Commissione formata da docenti universitari e bibliotecari conservatori fu approvata con Decreto ministeriale; dopo due anni il lavoro della Commissione giunse alla strutturazione della scheda di censimento chiamata Jemolo-Morelli, dai nomi delle due bibliotecarie che recepirono le decisioni della Commissione e la compilarono.

La scheda presentata allora è quasi in tutto simile a quella che usiamo ancora oggi: essa si componeva e si compone di tre sezioni, la prima dedicata alla descrizione esterna (elementi codicologici, storia del manoscritto) la seconda a quella interna (autori, titoli, incipit/explicit), la terza alla bibliografia e a eventuali riproduzioni del manoscritto; la scheda venne poi sperimentata da alcune biblioteche romane di conservazione.

¹ Per queste notizie si è preso spunto dalla relazione: *Working with Manus Online*, presentata da Lucia Negrini presso la Società Internazionale per lo Studio del Medioevo Latino (SISMEL) nel settembre del 2012: Lucia Negrini, *Working with Manus online: the Italian cataloguing standard and the national data-base with descriptions of manuscripts held in Italian libraries*, 28 settembre 2012, <<https://www.sismelfirenze.it/index.php/contatti/itemlist/user/378-superuser?start=200>>.

Nel 1990 furono prodotti due importanti strumenti: il software Manus 1.0, che operava su sistema DOS, e la nuova edizione della *Guida a una descrizione uniforme dei manoscritti e al loro censimento*², con una sezione dedicata all'inserimento dei dati, e con Appendici per la descrizione dei manoscritti decorati, di quelli con notazione musicale, dei manoscritti liturgici.

Il programma fu testato con l'inserimento delle schede descrittive dei manoscritti 701-800 della Biblioteca Casanatense.

Alla versione 1.0 seguirono le versioni 2.0 (1997) e 3.0 (2000); i programmi operavano in ambiente Windows, che facilitava notevolmente l'inserimento e la gestione dei dati; il software permetteva inoltre anche di lavorare in piccole reti locali. La scheda venne arricchita con la sezione dedicata ai carteggi e venne aggiunto il campo *Camicia*, destinato alla descrizione di manoscritti costituiti da materiale diverso (disegni, lettere, trascrizioni di documenti, appunti ecc.), cui si voglia dare un titolo d'insieme; inoltre, fu infine aggiunta una funzione che permetteva l'inserimento di immagini in ciascun campo.

A partire dal 2003 i dati raccolti, inviati all'ICCU dalle biblioteche su compact disc, poterono essere interrogati e visualizzati dall'utenza su un portale dedicato.

Un'altra importante realizzazione, agli inizi degli anni 2000, è stata la produzione del software Manus 4.0, utilizzato nell'ambito del progetto europeo Rinascimento Virtuale per la raccolta dei dati di palinsesti greci conservati nelle biblioteche delle nazioni europee partecipanti al progetto.

Nel frattempo, però, la tecnologia informatica progrediva, ed era oramai possibile pensare a una catalogazione online.

MOL 1.0 fu presentato il 10 luglio 2007 nel corso della giornata di studio: "La nuova versione di *Manus* per la catalogazione online", presso la Sala Conferenze della Biblioteca nazionale centrale di Roma, e venne accolta con grande interesse da bibliotecari e studiosi del settore.

3. L'Authority File e le Linee Guida

Un significativo risultato delle attività svolte dall'Area manoscritti dell'ICCU è stato la pubblicazione, nel 2018, delle *Linee Guida per la formulazione e il trattamento in Manus Online delle voci di autorità di nomi di persone, enti, famiglie e luoghi*, realizzate dal Gruppo di lavoro per la gestione e la manutenzione dell'Authority File di Manus Online, formato dalle bibliotecarie dell'Area manoscritti e dell'Area standard dell'ICCU, da colleghe e colleghi di altre biblioteche e da esperti del settore.

Questa brossura contiene le indicazioni metodologiche per «formulare in modo omogeneo le registrazioni di autorità in Manus Online», sulla base delle REICAT e delle norme per la registrazione dei nomi nell'Authority file di SBN.

² Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, *Guida a una descrizione uniforme dei manoscritti e al loro censimento*, a cura di V. Jemolo e M. Morelli, Roma: [s. n.], 1990.

In questa piccola pubblicazione (in totale 45 pagine) c'è un capitolo, a mio avviso, di estrema importanza per chi si accinga a catalogare un manoscritto: quello relativo ai nomi non identificati³.

Infatti, uno dei grossi problemi che si incontrano quando si descrive un manoscritto, soprattutto di epoca medievale, è che una quantità significativa di nomi presenti nel manoscritto non trovi riscontro al di fuori dello stesso. La scheda catalografica per la quale si crea una nuova voce di autorità risulta in tal caso essere l'unica fonte in grado di giustificare l'esistenza di quel nome specifico. Poiché alcuni nomi non sono identificabili attraverso repertori e altra bibliografia scientifica, nemmeno dopo una ricerca attenta e approfondita, tali nomi possono solo essere registrati in forma normalizzata all'interno di una lista di autorità, ma in seguito potrebbero essere modificati, ed eventualmente identificati.

4. 16 dicembre 2021: la nascita del portale Alfabetica e dell'ecosistema ICCU

A partire dal 2019 l'ICCU è stato impegnato nella completa reingegnerizzazione dei programmi e delle basi dati esistenti, create e gestite dall'ICCU stesso.

Proprio in virtù di queste profonde trasformazioni di tipo tecnico, MOL è ora in grado di gestire un numero molto maggiore di funzionalità, e si presenta molto più duttile, consentendo quindi anche la catalogazione di diverse tipologie di materiali manoscritti: liturgici, giuridici, di alfabeti e scritture diversi da quella latina ecc.⁴

Manus Online non è quindi più un programma a sé stante, ma è parte integrante del network del nostro Istituto, rendendo visibili i dati non solo dal suo front-end, ma anche dalle interrogazioni effettuate dall'OPAC SBN e da Alfabetica⁵. Una rilevante novità è poi la possibilità di collegare alla descrizione di un manoscritto le sue riproduzioni digitali integrali, sia dei manoscritti presenti nella Teca dell'ICCU che a quelle reperibili in fonti esterne; inoltre, all'interno di ciascuna scheda possono essere inserite singole immagini che, a giudizio del catalogatore, risultino essere di corredo ai dati descrittivi.

Nel back-end la scheda è stata arricchita di alcuni campi, e alcuni elementi già esistenti sono stati ulteriormente approfonditi; le novità presenti nella scheda non sono solo il frutto delle nostre riflessioni, ma anche dei consigli e suggerimenti di

³ Cfr. Gruppo di Lavoro per la gestione e la manutenzione dell'Authority File di Manus Online, *Linee Guida per la formulazione e il trattamento in Manus Online (MOL) delle voci di autorità di nomi di persone, di enti, di famiglie e di luoghi*, Roma: ICCU, 2018, p. 7-8, reperibile all'indirizzo: <<https://manus.iccu.sbn.it/norme-catalografiche2>>.

⁴ Cfr. Valentina Atturo – Lucia Negrini, *Manus Online: una base dati rinnovata, non solo per gli specialisti*, «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), 1, p. 76-91, <<https://digitalia.cultura.gov.it/article/view/2959/2058>>.

⁵ Per notizie su Alfabetica e sul nuovo OPAC di SBN cfr. i relativi contributi presenti in «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), 1, che contiene una sezione dedicata ad Alfabetica e al nuovo ecosistema dei servizi bibliografici nazionali.

catalogatori e studiosi di settori specialistici, in modo da renderla adatta per la descrizione di varie tipologie di manoscritti.

Ad esempio, il campo *Datazione* è stato arricchito di un menù a tendina con le diverse ere che si possono trovare attestate nei manoscritti conservati nelle nostre biblioteche: era cristiana, era bizantina (o della fondazione del mondo), era della fondazione di Roma, era di Diocleziano o dei martiri, Egira; segue un sottocampo dedicato alla *Formulazione della datazione*; nel campo *Decorazione*, nel sottocampo *Pagine ornate* è stato aggiunto un menù a tendina per evidenziare la presenza di *Unwan* o *Pyle*⁶.

Un campo assolutamente nuovo è quello dedicato alla segnalazione della *Pecia*: in vista della catalogazione di manoscritti giuridici (progetto speciale MOL Iuridica) e/o universitari, abbiamo pensato di creare *ex novo* questo campo, dove verranno inseriti i dati relativi alle carte e poi una trascrizione delle note di *pecia*.

Ci sono novità anche in relazione alla *Bibliografia*: la citazione bibliografica può essere compilata sia in modo non strutturato, cioè inserendo le citazioni in un campo libero, una di seguito all'altra, come nella versione precedente, sia in modo strutturato: la citazione sarà cioè catturabile da una base dati parallela; ciò comporterà sia un notevole risparmio di lavoro per i catalogatori, sia anche la certezza di usufruire di citazioni finalmente uniformi in tutte le schede descrittive, aggiungendo poi nella singola scheda la/le pagine di riferimento.

5. I Corsi su Manus Online. I tirocini. Le collaborazioni

A partire dal 2018 l'Area manoscritti dell'ICCU si è impegnata a organizzare corsi base sulla scheda e la procedura.

Sono stati organizzati corsi nella sede dell'ICCU, in cooperazione con le sezioni dell'Associazione italiana biblioteche (AIB) di tutte le regioni, con istituzioni e università: abbiamo calcolato che a questi corsi hanno partecipato complessivamente dalle 1.400 alle 1.500 persone; in tempo di pandemia i corsi si sono svolti da remoto.

In questi anni abbiamo avuto poi il piacere di ospitare presso di noi anche volontari e tirocinanti, che sono stati seguiti con attenzione dal personale dell'Area manoscritti, con la forte consapevolezza che il futuro è nelle mani delle nuove generazioni.

Last but not least, accenno ai progetti realizzati, o in fase di realizzazione, con questa e altre università, e alla stretta collaborazione con la Conferenza Episcopale Italiana (CEI), assieme alla quale, proprio in questi tempi, stiamo dando corpo alla realizzazione di una descrizione dei manoscritti liturgici su una *costola* della scheda Manus: il progetto, di comune accordo, è stato chiamato MOL Liturgica.

La solida base di Manus Online è il lavoro quotidiano delle biblioteche aderenti; il personale impiegato per la catalogazione, come tutti sappiamo, non è più costitui-

⁶ L'*Unwan* è un elemento decorativo posto a inizio pagina proprio per marcare l'inizio del testo ed è tipico dei manoscritti più sontuosi in scrittura araba; la *pyle* (*porta* in greco) è un elemento decorativo di forma rettangolare che incornicia la parte iniziale della colonna di scrittura.

to solo dai bibliotecari conservatori, che ora figurano di norma come coordinatori di progetto. Purtroppo nelle nostre biblioteche gli organici si sono impoveriti e generalmente il lavoro di catalogazione viene affidato a personale assunto *a progetto*: si tratta in massima parte di giovani laureati e di personale specializzato con *curricula* di tutto rispetto, in una situazione di precarietà che purtroppo nel nostro paese costituisce ormai la norma: speriamo questo panorama possa cambiare. Non sarebbe stato possibile raggiungere questi risultati senza un forte spirito di cooperazione fra tutti; senza un attento e paziente ascolto reciproco fra tutti i partner coinvolti, forti tutti della convinzione che nessun lavoro portato avanti individualmente avrebbe mai potuto farci arrivare ad un risultato di questa portata. Mi fermo qui; vorrei però terminare questo mio contributo con una citazione letteraria di alcuni brani, che mi sono permessa di “assemblare”, dall’ultimo capitolo del *Nome della Rosa*, che mi hanno sempre colpito per la loro forza espressiva e che sembrano descrivere icasticamente il paziente lavoro di ricostruzione di chi, con professionalità e umiltà, si accinge a descrivere questo materiale così particolare conservato nei nostri archivi e biblioteche:

«Rovistando tra le macerie trovavo a tratti brandelli di pergamena, precipitati dallo scriptorium e dalla biblioteca e sopravvissuti come tesori sepolti nella terra; e incominciai a raccogliarli, come se dovessi ricomporre i fogli di un libro [...] Lungo il viaggio di ritorno e poi a Melk passai molte ore a tentar di decifrare quelle vestigia. Spesso riconobbi da una parola o da una immagine residua di quale opera si trattasse. Quando ritrovai nel tempo altre copie di quei libri, li studiai con amore, come se il fato mi avesse lasciato quel legato, come se l’averne individuato la copia distrutta fosse stato un segno chiaro del cielo che diceva: tolle et lege. Alla fine della mia paziente ricomposizione mi si disegnò come una biblioteca minore, segno di quella maggiore scomparsa, una biblioteca fatta di brani, citazioni, periodi incompiuti, moncherini di libri»⁷.

6. I frammenti in MOL

L’esigenza di introdurre alcune modifiche orientate alla descrizione dei frammenti all’interno di MOL 2.0 è stata dettata dalle richieste, sempre più numerose da parte di biblioteche e archivi, di poter descrivere fondi interamente costituiti da frammenti di codici, nella quasi generalità dei casi distaccati da legature e conservati a parte. Oltre ai frammenti del progetto Textus invisibilis, per il quale è stata stipulata una specifica convenzione con l’Università degli Studi di Urbino, allo stato attuale sono coinvolti nella catalogazione di fondi di frammenti con Manus Online: l’Archivio di Stato di Roma, con il quale l’ICCU ha parimenti stipulato un’apposita convenzione; l’Archivio di Stato di Bologna, con 280 descrizioni all’attivo⁸ grazie all’opera di catalogazione della dott.ssa Rita De Tata; il progetto FRAC (Fragments

⁷ Umberto Eco, *Il nome della Rosa*, Milano: Bompiani, 1980, p. 471-472.

⁸ Dato desunto il 9 gennaio 2024.

of Commedia)⁹, che ha come obiettivo il censimento e la descrizione dei frammenti della Commedia dantesca; i frammenti dell'Archivio di Stato di Arezzo, a cura della dott.ssa Ilaria Marcelli; i frammenti della Biblioteca Multimediale di Fabriano "Romualdo Sassi", a cui sta lavorando la dott.ssa Patricia Bartocchetti¹⁰.

In realtà, già prima della reingegnerizzazione che ha portato a MOL 2.0 era possibile descrivere un frammento all'interno del database Manus, poiché la struttura estremamente flessibile del programma già consentiva di adattarlo a situazioni molto differenti l'una dall'altra. Tuttavia, un vistoso limite era rappresentato in primo luogo dall'impossibilità di ricercare i codici in stato di frammento all'interno del database e quindi di poterli enucleare come una delle tante possibili categorie in cui i manoscritti possono essere raggruppati. Sono state pertanto introdotte alcune modifiche (non moltissime) che ora consentono di considerare più nello specifico la realtà "frammento". D'altro canto non è stata avvertita la necessità di creare una maschera apposita per la descrizione dei soli frammenti, dal momento che la scelta ideale che sta a monte di MOL è che i frammenti non costituiscono una realtà altra e differente rispetto al codice nella sua interezza. Se si rinuncia a inserire il frammento, per quanto piccolo e malridotto possa essere, in un contesto più ampio che è quello del codice da cui proviene e di cui forse resta l'ultima traccia esistente, si perde la capacità di comprensione profonda di quel documento storico. È questa una convinzione profondamente radicata nel gruppo di lavoro dell'Area manoscritti dell'ICCU, pur nella consapevolezza che il modo di accostarsi a un frammento richiede comunque alcune accortezze speciali, necessarie a fornire tutti quegli indizi che ci permettono, a volte di restituire, a volte solo di tratteggiare molto pallidamente l'aspetto di un *intero* altrimenti perduto: ma anche in questo caso, solo l'esperienza di chi ha avuto una formazione specialistica sugli "interi" consente di meglio contestualizzare i *membra disiecta* all'interno di un realtà non più esistente. Tutto questo viene detto non per alimentare sterili contrapposizioni tra questa posizione e quella dei sostenitori della *fragmentologia*¹¹, dal momento che ogni posizione, se poi finalizzata all'ampliamento delle conoscenze umane, è rispettabile e accettabile, ma per spiegare in modo chiaro da quale presupposto teorico si è partiti per il rinnovamento di MOL. Allo stato attuale Manus Online 2.0 permette di descrivere un codice, a prescindere da quali siano le condizioni in cui esso è giunto

⁹ <https://manus.iccu.sbn.it/frammenti-della-commedia-frac>.

¹⁰ Per alcune riflessioni generali sul coinvolgimento dell'ICCU nella catalogazione dei frammenti v. Simonetta Buttò, *Il programma Manus e la catalogazione di frammenti di codici in Italia*, in: *Frammenti di un discorso storico. Per una grammatica dell'aldilà del frammento*, a cura di C. Tristano, Spoleto: Centro italiano di studi sull'alto medioevo, 2019, p. 473-479: Si veda inoltre: Alessandra Corbo, *Manus e la catalogazione informatizzata del frammento manoscritto: riflessioni e proposte*, «Archivi», 13 (2018), n. 1, p. 29-42.

¹¹ Si veda l'interessante editoriale a firma di William Duba e Christoph Flüeler, *Fragments and fragmentology*, «Fragmentology», 1 (2018), p. 1-5, scritto per presentare al pubblico per la prima volta la rivista dal titolo programmatico, consultabile all'indirizzo: <<https://www.fragmentology.ms/article/view/editorial/2713>>.

fino a noi: che il codice sia perfettamente integro, o acefalo, mutilo o lacunoso, o che di esso sia rimasta una sola carta, mezza carta, finanche una porzione delle dimensioni di un francobollo, si tratta sempre di un manoscritto e come tale viene valutato all'interno del nostro database.

In questa prospettiva, non era dunque possibile separare la banca dati in due tronconi: gli interi e i frammenti, anche perché ciò avrebbe inciso sulla ricercabilità dei dati, quando la ricerca fosse di tipo trasversale: se si ha bisogno di sapere, per esempio, quanti e quali manoscritti presenti in Manus tramandano le Commedie di Terenzio, sarà fondamentale per il ricercatore recuperarne tutti i testimoni, anche se alcuni sono ridotti a frustuli con qualche verso, poiché sono pur sempre l'ineludibile testimonianza di esemplari un tempo esistiti e probabilmente letti.

Figura 1. Immagine del campo *Composizione materiale*

Se tuttavia la finalità dell'utente è quella, al contrario, di individuare in primo luogo ciò che è frammento, la principale implementazione apportata a MOL ha riguardato, nella scheda di descrizione del back-end, il campo *Composizione materiale* (Fig. 1): si tratta della seconda pagina in cui si articola la Descrizione esterna ed è quella che consente di individuare da subito la natura dell'oggetto che si va a descrivere (se cioè il codice è unitario, composito, ha altra forma - come rotolo o altro - oppure è un frammento). Evidenziando questo semplice flag è possibile procedere all'indicizzazione automatica dell'oggetto nella sua natura di frammento, facendo sì che sulla scheda visualizzata dall'utente compaia la voce "Frammento" accanto alla descrizione segnaletica e a quella per esteso (Fig. 2-3).

Figura 2. Scheda segnaletica del frammento (front-end)



Figura 3. Scheda estesa del frammento (front-end)

Una volta dichiarato che quanto si va a descrivere è, appunto, un frammento, il resto della scheda rimane inalterato, dal momento che in Manus è possibile riempire solo i campi che possono interessare il singolo oggetto (o quelli che il catalogatore sceglierà di valorizzare) (Fig. 4): i soli campi obbligatori, senza i quali non è possi-

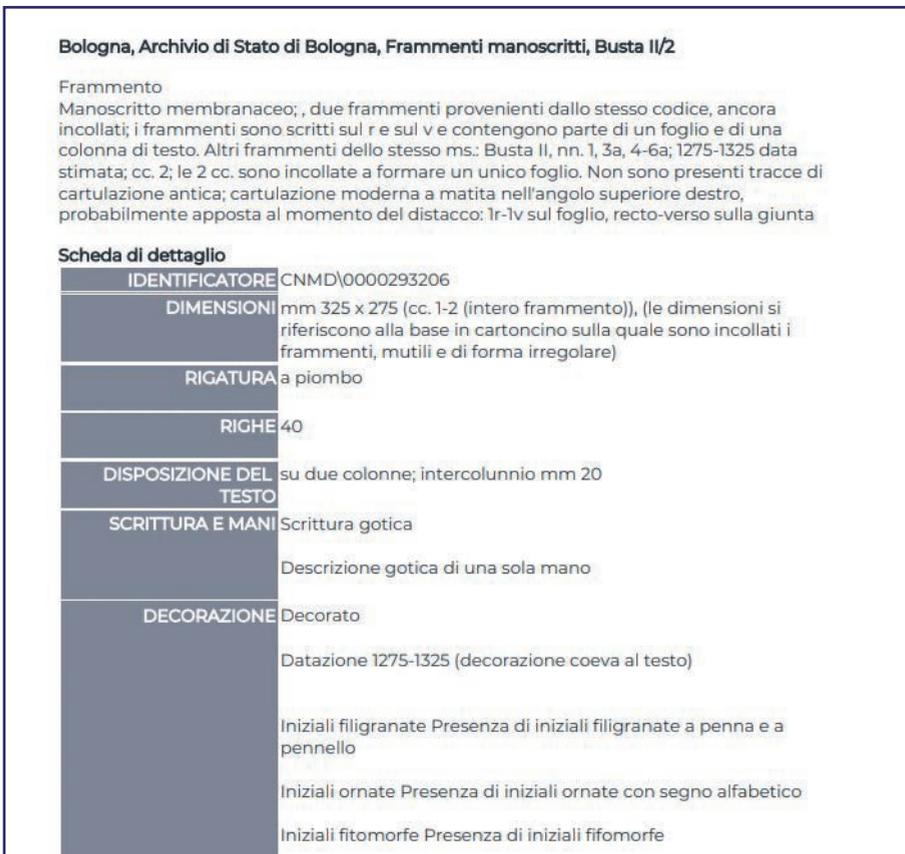


Figura 4. Modello di scheda di frammento descritto in MOL 2.0 (front-end)

bile pubblicare la scheda (ma ciò riguarda qualsiasi testimonianza scritta la cui descrizione è accolta in MOL), sono la datazione, la materia del supporto, la consistenza e le dimensioni, quattro dati fondamentali che riguardano l'intero come il parziale. Il corredo delle immagini è un altro punto essenziale del nuovo Manus: esso ora offre la possibilità di allegare molte immagini per ogni singola scheda, compresi i particolari, così come consente all'utente di valersi del IIIF (International Image Interoperability Framework) (Fig. 5), estremamente utile per gli studiosi che vogliono mettere a confronto immagini di codici e ancora di più di frammenti coerenti tra loro, ma conservati in luoghi diversi. Ciò risulta appunto particolarmente efficace per i frammenti, la cui riproduzione integrale richiede per altro solitamente poco sforzo.

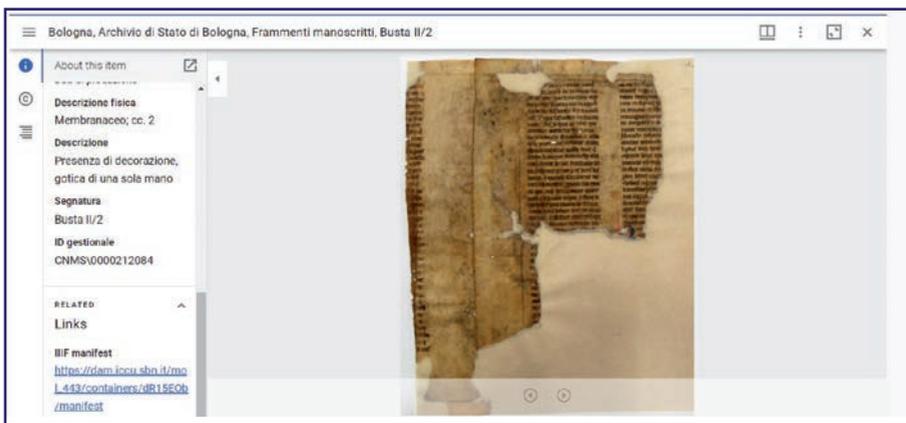


Figura 5. Immagine di frammento in MOL 2.0

Quanto detto sinora riguarda il caso in cui si intenda descrivere un frammento come unità a sé stante, per la quale viene creata un'apposita scheda. Prima di procedere oltre, occorre fare a questo punto due considerazioni che sono il frutto di una riflessione avvenuta tempo fa, in occasione di un incontro tenutosi presso la Sala riunioni dell'ICCU il 3 novembre 2022 e a cui hanno partecipato catalogatori e studiosi che si occupano di frammenti. In primo luogo, una domanda, apparentemente banale, ma, come si dirà, di fatto sostanziale, riguarda la definizione dell'oggetto: cosa intendiamo noi con frammento? Questa domanda, lanciata nella nostra discussione da Caterina Tristano, ci ha lasciato in un primo momento spiazzati. Tutti, infatti, penso sono concordi nel ritenere alcune *entità ex libris* come frammenti: la porzione di una carta, un'intera carta, più o meno menomata, un bifolio rientrano nell'accezione comune di frammento. Ma fino a che punto possiamo, e vogliamo soprattutto, estendere il concetto di frammento? Qui non ci sono opinioni più o meno valide, si tratta di fare una scelta a monte, che coinvolgerà tutti coloro che decideranno di utilizzare MOL 2.0 come strumento per la catalogazione dei propri fondi manoscritti. Al momento ciò su cui abbiamo concordato è

che in Manus sarà considerato frammento tutto ciò che ha un'estensione non superiore ad un fascicolo. Pertanto, se dovessimo descrivere oggi in Manus il famosissimo *Virgilio Augusteo* (Vat. lat. 3256)¹² lo denomineremmo "frammento", mentre il manoscritto Roma, Biblioteca Vallicelliana F 25/8¹³, un insieme di bifolia provenienti da diversi fascicoli di un codice del secolo XII con le *Metamorfosi* di Ovidio lo considereremo codice unitario, secondo quella che è ovviamente, è bene ribadirlo, una pura convenzione.

Un'altra questione che si è posta riguarda la segnatura di un frammento che, ancora *in situ*, si voglia descrivere in una scheda autonoma: come mantenere il legame con il suo *contenitore*? Per risolvere tale problema Manus offre due possibilità: o considerare il manoscritto come un composito e trattare il frammento come una delle unità che lo costituiscono (questa scelta è però sconsigliata, perché la presenza di un frammento all'interno di un codice non rientra tecnicamente nel concetto di composito), oppure creare una scheda che abbia come segnatura la stessa del codice con l'aggiunta di *bis, ter* ecc. oppure *a, b, c* e simili. In questo modo è possibile mantenere l'unità di segnatura, pur descrivendo due oggetti completamente differenti. Nella *Storia del codice* dovrà essere esplicitato il particolare rapporto che lega l'unità minore (frustulo) all'unità maggiore (il codice).

In ogni caso, ciò che è apparso a tutti noi condivisibile è il fatto che ogni frammento vada descritto separatamente, fatta eccezione per due o più elementi ritrovati nel medesimo contesto e solidali tra loro: questo vale anche nel caso in cui si posseggono frammenti riconducibili senza ombra di dubbio a un medesimo codi-

DESCRIZIONE DELLA SEGNATURA	
Denominazione segnatura	Cass. 277, nr. 7
Collocazione fisica	
Alias	
CNMS	CNMS\0000313843
Biblioteca	Archivio di Stato di Roma
Fondo	Collezione delle pergamene
Lingue utilizzate	
Alfabeti	
Numeri di Inventario	
Raggruppamento segnature	
Raggruppamento segnature	

Figura 6. *Raggruppamento segnature (back-end)*

¹² Il codice è consultabile all'indirizzo: <https://digi.vatlib.it/view/MSS_Vat.lat.3256>.

¹³ Su questo codice si veda Elisabetta Caldelli, *I manoscritti classici latini della Biblioteca Vallicelliana di Roma*, Roma: Istituto Poligrafico e Zecca dello Stato, 2021, p. 112-113.

ce, ma provenienti da punti diversi dello stesso, ritrovati in contesti differenti e magari anche provvisti ciascuno di una propria segnatura (situazione molto frequente nel caso di raccolte di frammenti staccati).

Ricordo a tale proposito che Manus è stato concepito non per studiare i frammenti al fine di offrire una proposta ricostruttiva dell'intero (operazione che spetta allo studioso), ma al fine di catalogarli, cioè al fine di individuare e fornire dati identificanti dell'oggetto che si va a descrivere, utili a chi poi li dovrà studiare. Tuttavia, per consentire al catalogatore di collegare due o più frammenti evidentemente provenienti dal medesimo codice (o che la bibliografia esistente indica come provenienti da uno stesso manoscritto) è stata introdotta, in *Descrizione della segnatura*, la funzione *Raggruppamento segnature* (Fig. 6), che ci consente di raggruppare nell'ambito di una ricerca in front-end tutte quelle segnature che sono, per varie ragioni, legate tra di loro.

Quanto detto riguarda una scheda di descrizione di un frammento. In Manus, però, esiste una casistica molto ampia di modalità di approccio al frammento che non possono essere ignorate: è frequente, infatti, che il catalogatore che sta descrivendo un codice all'interno del quale sia presente un frammento, non abbia la possibilità di catalogarlo (per mancanza di tempo, di competenze ecc.). In questo caso, può limitarsi a segnalarne l'esistenza all'interno della scheda relativa al codice: in *Descrizione esterna* (D.E.) nel campo *Consistenza* troverà una sezione dedicata ai frammenti *in situ* (Fig. 7), in cui può dare conto della presenza del frammento indicandone la posizione e, usufruendo del menù a tendina (Fig. 8) realizzato come

The image shows a screenshot of a web-based form titled "8. CONSISTENZA". The form is organized into several sections:

- Numero carte:** A label followed by a text input field.
- Guardie iniziali:** A label followed by a text input field.
- Corpo:** A label followed by a text input field containing the number "1".
- Guardie finali:** A label followed by a text input field.
- Mancanze:** A label followed by a large text area with a small icon in the bottom right corner.
- Parti a stampa:** A label followed by a large text area with a small icon in the bottom right corner.
- Frammenti in situ:** A sub-section header followed by:
 - Posizione:** A label followed by a dropdown menu.
 - Carta:** A label followed by a dropdown menu.
 - Note alla sezione 8:** A label followed by a large text area with a small icon in the bottom right corner.

Figura 7. Il campo *Consistenza* (back-end)

implementabile, la tipologia del riuso. Se poi il catalogatore vuole dare qualche informazione in più sul frammento, nella pagina *Elementi paleografico codicologici* (pagina a scorrimento verticale), troverà anche una sezione denominata *Frammenti*, un campo libero in cui potrà fornire, in forma discorsiva, tutte le informazioni che riterrà utili per il lettore, dare conto schematicamente della porzione di carta/bifolio da cui il frammento proviene, cliccando sull'allegato *Schema di descrizione del frammento* (Fig. 9), o anche allegare semplicemente un'immagine.

Frammenti in situ

Posizione

Carta

Note alla sezione 8

- Brachetta
- Carta di guardia
- Legatura
- Offset
- Rinforzo del dorso

Figura 8. Menù a tendina nel campo Consistenza

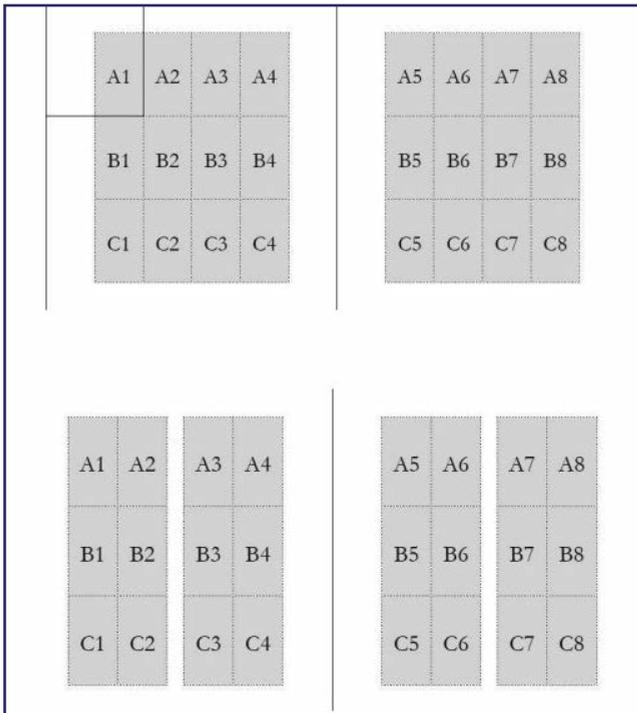


Figura 9. Schema di descrizione del frammento

Allo stato attuale queste sono le possibilità che offre Manus per la descrizione/segnalazione dei frammenti. Il cammino però è ben lungi dall'essere concluso, né un sistema predefinito può pretendere di coprire l'infinita casistica nella quale ci si può imbattere quando si ha a che fare con il riutilizzo dei codici, a volte veramente fantasioso. Si pensi, ad esempio, ai cosiddetti corali di Penne, scoperti da Carlo Tedeschi nell'Archivio storico dell'Arcidiocesi di Pescara e Penne e ora in fase di studio¹⁴, in cui i restauri antichi del codice principale con pezzi di altri codici, hanno prodotto giustapposizioni di un frustulo sull'altro: in questi casi si comprende bene quanto sia altissima la difficoltà di descriverli seguendo una maschera preconstituita senza tradire la natura stessa dell'oggetto.

A conclusione di questo breve *excursus* sul frammento in Manus, è opportuno riallacciarsi ancora una volta alle conclusioni e alle proposte avanzate durante l'incontro del 3 novembre sopra ricordato: la catalogazione dei frammenti pone l'operatore di fronte a scelte e dunque a dilemmi, rispetto ai quali non può e non deve sentirsi solo, poiché uno dei valori fondanti di Manus Online è proprio il senso di comunità, basata sul confronto e sullo scambio. Certamente sarà cura dell'Area manoscritti dettare le norme generali per la catalogazione dei frammenti, ma sarà soprattutto la condivisione delle esperienze a dare nuovi apporti al programma. In questo senso si è parlato di organizzare seminari periodici, che permettano di raffinare l'impostazione generale di Manus, ma che possano anche offrire nuovi spunti e aprire nuove frontiere.

In the last two years Manus Online (MOL) was subjected to relevant re-engineering activities. The new version - Manus Online 2.0 - was published in December 2021 and it allowed to introduce new important features, among which the possibility to catalogue and index manuscript fragments gathered in public, ecclesiastical and private libraries and archives. In this contribution the theoretical assumptions and technical aspects of the MOL back-end relating to the description of fragments in MOL will be outlined.

¹⁴ Carlo Tedeschi – Ilaria Vezzosi, *Una biblioteca nascosta. Frammenti di una Bibbia atlantica e di altri codici nell'archivio storico diocesano di Penne e Pescara*, in: *Frammenti di un discorso storico. Per una grammatica dell'aldilà del frammento*, a cura di C. Tristano, Spoleto: CISAM, 2019, p. 561-586.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024.

IVS Commune *online*: proposta di un sistema integrato per lo studio della trasmissione dei testi (del diritto)

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00097

Alessandra Panzanelli
Università di Torino

“Towards a corpus of texts of the ius commune from manuscript to print”: così recita il sottotitolo di IVS Commune online, lo strumento digitale cui è dedicato il presente articolo. Tale sottotitolo chiarisce, o dovrebbe chiarire, l’obiettivo principale di questo “nuovo” strumento digitale, che nasce per facilitare la ricostruzione di un corpus di testi che, notoriamente e per sua natura, si presenta ricco, complesso e molto difficile da governare. Un paragrafo iniziale serve a delineare i confini del corpus. Definito l’oggetto della ricerca, nonché il contesto accademico in cui essa è nata, l’articolo si sviluppa nella illustrazione dello strumento, delle logiche ad esso sottese, delle principali funzionalità, insistendo particolarmente sugli aspetti legati alla interoperabilità, al riuso dei dati e, da ultimo ma non per importanza, alla cooperazione: tra editors, categoria che riunisce coloro che usano, popolandola, la base dati, lavorando in contesti diversi – accademia e istituti di conservazione. Il contributo si chiude con un cenno ai più recenti sviluppi dello strumento e alle prospettive future, di breve e lungo periodo.

1. L’oggetto e il progetto

IVS Commune *online* (in breve ICo)¹ è uno strumento digitale sviluppato per facilitare lo studio della trasmissione dei testi del diritto sul lungo periodo, dal medioevo all’età moderna, periodo in cui un momento cruciale è naturalmente costituito dalla introduzione della stampa a caratteri mobili, non a caso sottolineato nel titolo completo della base dati. La “rivoluzione della stampa” tuttavia non

¹ <<https://www.iuscommuneonline.unito.it>>. La base dati è uno dei risultati principali del progetto di ricerca “Le fonti del diritto nell’Europa medievale e moderna. Per un corpus dei testi dello *ius commune* dal manoscritto alla stampa (1350-1650)”, premiato dal Programma Giovani Ricercatori “Rita Levi Montalcini” (bando 2017) che ne ha supportato la prima fase di realizzazione. Dal dicembre 2023 il progetto conosce una seconda fase di sviluppo finanziata nell’ambito dello Spoke 3 (Digital Libraries, Archives and Philology) del partenariato esteso CHANGES Cultural Heritage Active Innovation for Nex-Gen Sustainable Society Extended Partnership (Progetto PE 0000020 CHANGES, - CUP D53C22002530006, PNRR Missione 4 Componente 2 Investimento 1.3, finanziato dall’Unione europea – NextGenerationEU). Quanto alla scelta del nome IVS Commune *online*, questa rinvia al concetto di *ius commune* come elaborato all’indomani del secondo conflitto mondiale quando un sistema di diritto condiviso a livello continentale sembrò configurare un precedente per

esaurisce le domande per rispondere alle quali ICo mira a raccogliere in un unico ambiente tutti i dati e le informazioni rilevanti. Il cuore di ICo è una base dati bibliografica, concepita per integrarsi al meglio con i principali strumenti e progetti che insistono, in tutto o in parte, sullo stesso terreno. La perfetta integrazione (intesa come interoperabilità) è l'obiettivo ideale, al quale ci si approssima in vario modo, a seconda delle tecnologie con cui le altre risorse sono sviluppate: si va dal semplice link all'allestimento di forme di interoperabilità tra sistemi. Nel mezzo c'è il lavoro di editor che contribuiscono a più di un progetto. L'integrazione con altri sistemi mira a sua volta a creare un ecosistema, il cui focus sono i testi giuridici, ma che stimola a ragionare sulle tante categorie che informano in generale l'universo bibliografico (opere, testi, edizioni, esemplari) le quali emergono sotto una luce diversa, sia per le specificità dell'ambito disciplinare sia anche per il taglio di lungo periodo e trasversale (manoscritti, antiche edizioni, edizioni moderne). I concetti di edito e inedito, pubblicato e non pubblicato, diventano al contempo più e meno chiari; la distinzione tra manoscritto e stampato si assottiglia, mentre acquistano importanza gli ambienti di produzione e d'uso. Per tenere insieme tutto, senza perdere le specificità necessarie alla corretta comprensione dei fenomeni che si vogliono studiare, la base dati prevede appunto legami con tutte le risorse rilevanti e pertinenti in modo da creare un ambiente di studio specializzato ma che tiene conto del contesto bibliografico più ampio².

la costruzione di una comune cornice giuridica europea. *The common legal past of Europe* recita, con grande efficacia, la traduzione inglese di Manlio Bellomo, *L'Europa del diritto comune*, Roma: Il Cigno Galileo Galilei, 1988 (trad. a cura di Lydia Cochrane, 1995). Ad una visione che vedeva lo *ius commune* come il diritto còlto, espressione delle massime autorità in opposizione ai diritti particolari di comunità (*iura propria*), si è sostituita una visione inclusiva. Una utile sintesi è offerta da Alain Wijffels: «Le droit romain et le droit canonique constituent ce que l'historiographie désigne de "droits savants" (en opposition aux coutumes et législations particulières, qui longtemps, sauf exception, ne furent pas pleinement intégrées dans la science juridique), mais l'expression qui a été peut-être la plus propagée à notre époque est celle de *ius commune*. [...] En fait, *ius commune* était davantage un terme technique du langage juridique médiéval qui répondait au besoin, dans la société de l'époque, de maîtriser la multi-normativité». (Alain Wijffels, *Une très brève histoire du droit dans la civilisation occidentale (1000–2000)*, «Annales de Droit de Louvain», 77 (2017=2019), n. 3, p. 397–411: 399). Si veda quanto scrive Emanuele Conte che definisce lo *ius commune* come: «the complex of normative authorities and doctrinal interpretations produced by jurists from the 12th to the 15th century» chiosando «This relationship was not as conflictual as some of the legal historical literature depicts. Some examples [...] show how complicated the interwining of the *ius commune*, customary laws and municipal statutes in the late Middle Ages can be» (Emanuele Conte, *Consuetudine, Coutume, Gewohnheit and Ius Commune. An Introduction*, «Rechtsgeschichte - Legal History», 24 (2016), p. 234–243, cit. dall'abstract).

² Particolarmente felice è la circostanza per cui questo contributo appare nel numero seguente a quello che ospita *Fare per non sprecare. Nei laboratori del riuso digitale*. Atti del Convegno di studi (Roma, 9–10 giugno 2022), «DigItalia. Rivista del digitale nei beni culturali», 18 (2023), n. 2, p. 11–218. Il riuso, inteso anche come valorizzazione delle risorse ricevute in eredità, è un valore fondamentale per chi scrive e che informa il progetto ICo nella sua stessa concezione. È altresì importante poter illustrare IVS Commune *online* in questa sede, che appare come la più adeguata a proporre considerazioni che pertengono questioni squisitamente bibliografiche in ambiente digitale che non potevano essere illu-

Per questa ragione, pur consistendo anzitutto in una base di dati, ICo è talvolta presentata come piattaforma digitale; caratterizzata da una serie di elementi che la rendono diversa e – si spera – utile complemento agli strumenti già esistenti; illustrati e commentati nelle pagine che seguono, tali elementi si possono così sintetizzare:

- il focus della base dati è sul testo e sulla possibilità di seguirne le evoluzioni sul lungo periodo; concepita per ospitare descrizioni (e per identificare) ogni tipologia di testo: manoscritti e a stampa, documenti singoli o libri veri e propri;
- ciò non significa un misconoscimento delle peculiarità materiali del libro, cui anzi è dedicata un'ampia sezione;
- i dati sono corredati da attributi che ne esaltano il valore in fase di analisi critica: ad esempio i dati di incertezza circa la datazione o le attribuzioni sono esplicitati (non solo segnalati dall'uso della punteggiatura);
- dato l'obiettivo che si pone, ICo è progettata per agevolare la ricerca: la registrazione dei dati è impostata in modo che sia intuitiva anche per chi non ha grande familiarità con le categorie proprie del lavoro bibliografico;
- ICo mira ad integrarsi e allinearsi con i principali cataloghi generali così come si integra con gli altri strumenti digitali dedicati, ad alcuni dei quali contribuisce direttamente (attraverso forme di cooperazione, individuale e tra progetti).

Le domande di fondo a cui ICo mira a facilitare le risposte: quando, quanto e come un testo è mutato nei diversi passaggi, prima nella tradizione manoscritta e poi in quella a stampa? La scelta di campo, ovvero l'idea di esplorare un ambito specifico come quello giuridico, si deve in parte alla sfida che questo ambito comporta, per una tradizione testuale particolarmente complicata e complessa, da cui è derivata anche una frequentazione per così dire sincopata del libro giuridico da parte di bibliografi e storici del libro, che spesso hanno lasciato il campo alle cure di specialisti della disciplina³. È bene tuttavia chiarire subito che l'ambito giuridico non con-

strate nel dettaglio nelle altre sedi in cui ICo è stata presentata in relazione a specifiche piste di ricerca (Maria Alessandra Panzanelli Fratoni, *Bartolo da Sassoferrato: dal cuore d'Italia all'Europa. La fortuna del giurista tra antichi censimenti e nuovi progetti*, in: *I bestsellers del Medioevo e della prima Età Moderna. Tra evasione, studio e devozione*, a cura di E. Caldelli, Roma: Edizioni Efestò, 2023, p. 29-44); una presentazione sintetica si legge in Cristina Dondi – Matilde Malaspina, *L'ecosistema digitale del CERL per lo studio del libro antico a stampa: dal progetto 15cBOOKTRADE a oggi*, «*Digitalia. Rivista del digitale nei beni culturali*» 17 (2022), 1, p. 134-156: 146-147.

³ Studi e cataloghi fondamentali si devono a specialisti della disciplina, a partire dal censimento dei manoscritti di diritto romano, pionieristicamente realizzato con strumenti informatici da Gero Dolezalek (Gero Dolezalek, *Verzeichnis der Handschriften zum römischen Recht bis 1600*, Frankfurt am Main: Max Planck Institute, 1972) e da cui è nata poi *Manuscripta juridica* (<<https://manuscripts.rg.mpg.de/>>); una équipe di storici del diritto è responsabile del monumentale Catalogo dei codici del Collegio di Spagna (*I codici del Collegio di Spagna di Bologna*, studiati e descritti da Domenico Maffei et al., con la collaborazione di Mario Ascheri et al., Milano: Giuffrè, 1992), in buona parte ma non esclusivamente di diritto, anche quello base per uno sviluppo digitale (Progetto Irnerio: <<http://irnerio.cirfid.unibo.it/>>). A latere si dovrà notare come, a supporto delle loro ricerche, un certo numero di collezioni specializzate sono state create da studiosi, parte

diziona il cuore del modello concettuale sotteso alla base dati, stante che le categorie bibliografiche sono potenzialmente applicabili ad ogni ambito disciplinare. Quello giuridico, d’altro canto, presenta problematiche tali da configurarsi come un ottimo caso di studio, una palestra in cui affrontare questioni spinose (sulle attribuzioni ad esempio) e per sperimentare soluzioni da proporre poi anche altrove. Il molto lavoro già fatto (repertori, studi dedicati, cataloghi di singole collezioni), e i lavori che si stanno facendo da parte di singoli studiosi, costituiscono un punto di partenza molto avanzato. Perché questo “capitale” sia valorizzato, c’è bisogno di un luogo in cui possano confluire le diverse ricerche, passate e presenti, operazione che di per sé produce risultati⁴.

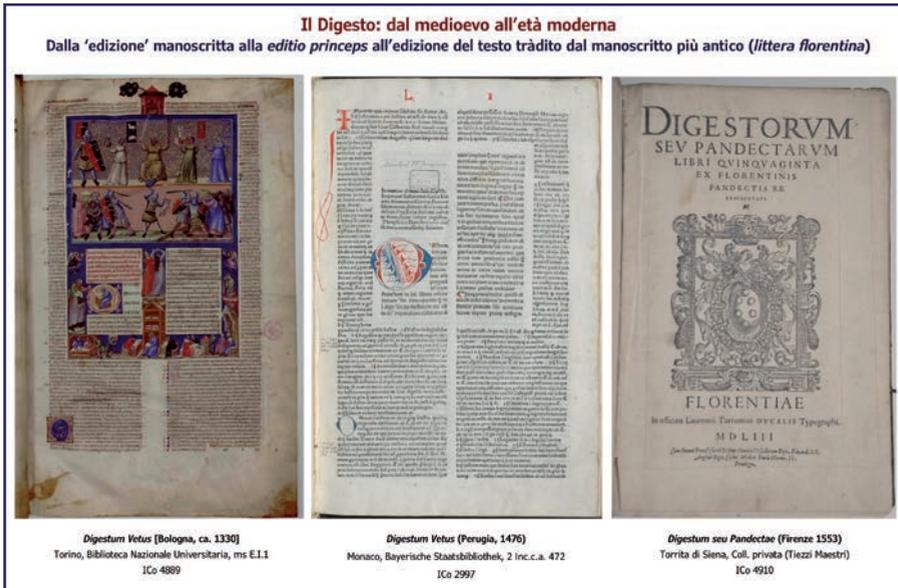


Figura 1. Le “pagine” simbolo della evoluzione dei testi scelte per l’home-page

delle quali forma oggi il prezioso nucleo di biblioteche istituzionali, si pensi ai Fondi Patetta (in Biblioteca Vaticana e nell’attuale Biblioteca Bobbio dell’Università di Torino) o al fondo Cortese oggi presso la Biblioteca del Senato; serve ancora allo scopo restando in famiglia la straordinaria raccolta di Domenico Maffei. Sulla necessità di affidarsi agli specialisti per dare conto correttamente delle pubblicazioni giuridiche già Gesner si esprime, nel libro XIX dei *Pandectae* 19, dedicato al diritto (Maria Alessandra Panzanelli Fratoni, *La biblioteca giuridica nella prima età moderna. Con un’indagine delle giurisprudenzia nella Grande Galleria di Carlo Emanuele I di Savoia*, in: *Reimmaginare la Grande Galleria: forme del sapere tra età moderna e culture digitali*, a cura di E. Guadagnin, F. Varallo e M. Vivarelli, Torino: Accademia University Press, 2022, p. 281–308: 286–287).

⁴ Lo dimostrano progetti come lo Incunabula Short-Title Catalogue che, riunendo in un unico punto le descrizioni delle edizioni incunabile ha trasformato un catalogo in un potente strumento di ricerca; Paul Needham, *ISTC as a tool for analytical bibliography*, in: *Bibliography and the study of 15th-century civilisation*, ed. by L. Hellinga and J. Goldfinch, London, 1987 (The British Library Occasional Papers, 5), p. 39–54; John Goldfinch — Karen Limper-Herz, *The Incunabula Short Title Catalogue (ISTC): Past, Present and Future*, in: *Printing R-evolution and Society. Fifty years that changed Europe (1450-1500)*, ed. by C. Dondi, Venezia: Edizioni Ca’ Foscari, 2020, p. 898-909.

Ricondurre *ad unum* è dunque idea chiave del progetto ICo, che ha appunto come primo obiettivo ospitare nello stesso luogo le descrizioni dei testi tràditi da manoscritti e di quelli pubblicati a stampa, al fine di facilitarne analisi e lettura comparate. Tre immagini sono state scelte per l'home-page del sito web di IVS Commune *online*, perché accostate l'una all'altra offrono una rappresentazione palpabile delle trasformazioni cui andò incontro il libro giuridico tra medioevo ed età moderna (Fig. 1). Si tratta delle prime pagine di tre diverse "edizioni" del Digesto, testo che bene rappresenta la tradizione testuale giuridica e gli studi di diritto. Una prima pagina proviene da un magnifico codice manoscritto di fattura bolognese, prodotto dunque nella patria del diritto medievale e databile alla prima metà del Trecento⁵; reca l'incipit del *Digestum Vetus* (la prima parte, secondo le partizioni tradizionali) disposto sulla pagina secondo lo schema canonizzato nelle università medievali, il testo autorevole incorniciato dalla glossa; una ricca miniatura integra la narrazione con complesse scene di giustizia, chiudendosi in basso con una rappresentazione vivacissima proprio della produzione libraria, dalla scrittura della pagina alla distribuzione dei libri, nel banco degli *stationarii*, fino alla loro collocazione sullo scaffale. A questa pagina, perfetta "fotografia" di una lunga fase nella storia del libro, si affianca la pagina d'incipit della *princeps*⁶, che fu messa in forma un secolo e mezzo più tardi da un gruppo di studenti; priva di miniature e decorata con semplicità, è sostanzialmente identica nei testi e nella mise-en-page. Chiude la triade il frontespizio di una edizione celeberrima: quella del testo del Digesto tràdito dal manoscritto più antico (il *Codex Florentinus*, già *Pisanus*, VI secolo), monumentale testimone di un'altra tradizione del testo (completo dunque dei cinquanta libri ma privo della glossa)⁷; l'edizione veniva a compiere un'impresa di lunghissimo corso (sulla *Littera Florentina* aveva speso grandi energie anche Angelo Poliziano e, tra i contemporanei, lo storico e giurista spagnolo Antonio Agustín) e fu portata a compimento nel 1553 in un contesto di grande ufficialità. Curata da Lelio Torelli, segretario del duca, e da suo figlio Francesco, l'edizione fu stampata da Lorenzo

⁵ Si tratta del ms E.1.1 della Biblioteca Nazionale Universitaria di Torino, celebre tra gli storici della miniatura per un apparato figurativo e decorativo di qualità altissima ricondotto alla mano del Maestro del 1328 (Massimo Medica, *Libri, miniatori e committenti nella Bologna di Bertrando del Poggetto*, in: *Giotto e le arti a Bologna al tempo di Bertrando del Poggetto (Bologna, Museo Civico Medievale, 3 dicembre 2005 - 28 marzo 2006)*, Cinisello Balsamo (Milano): Silvana Editoriale, 2005, p. 79-93).

⁶ Stampata a Perugia nel 1476, la *princeps* risulta identica nell'impaginato al manoscritto, ma è preceduta da una breve prefazione ricchissima di informazioni (Alessandra Panzanelli, *La stampa a Perugia nel Rinascimento: dai tipografi tedeschi agli editori locali*, Milano: FrancoAngeli, 2020, p. 65-72, 220).

⁷ Uno studio puntuale del codice è in Davide Baldi, *Il Codex Florentinus del Digesto e il "Fondo Pandette" della Biblioteca Laurenziana (con un'appendice di documenti inediti)**, «Segno e Testo» 8 (2010), p. 99-186, 10 tav.; sulla tradizione altomedievale dei testi del corpus giustiniano: Charles M. Radding — Antonio Ciaralli, *The Corpus Juris Civilis in the Middle Ages: Manuscripts and Transmission from the Sixth Century to the Juristic Revival*, Leiden: Brill, 2007, sulla datazione del *Florentinus* p. 36-40.

Torrentino, direttore della tipografia ducale. A suggellarne il valore di impresa pubblica è il grande stemma coronato dei Medici che campeggia sul frontespizio e un profluvio di privilegi, ottenuti dal papa, dall'imperatore e dai re di Francia e Inghilterra, oltre che ovviamente da Cosimo de' Medici duca di Firenze.

Miniature, impaginato, paratesti, presi insieme tutti questi elementi formano un cospicuo nucleo informativo; la base dati è progettata appunto per consentire una registrazione completa ma anche in parte intuitiva o comunque non cervellotica di ognuno e tutti i dati utili a studiare l'evoluzione dei testi. Ciò significa certamente testi e paratesti (e paratesti eccellenti possono essere raccolti nei frontespizi, come l'edizione delle Pandette fiorentine dimostra chiaramente); ma ciò significa anche tutti gli elementi pertinenti la struttura fisica del libro: collazione, misure, sistemi di scrittura o tipografia, presenza di apparati decorativi o iconografici, insomma tutti elementi pertinenti l'analisi codicologica e bibliologica.

Il risultato di tali analisi viene quindi presentato in modo che sia possibile collazionare tutti i dati raccolti, con le descrizioni affiancate le une alle altre (Fig. 2).

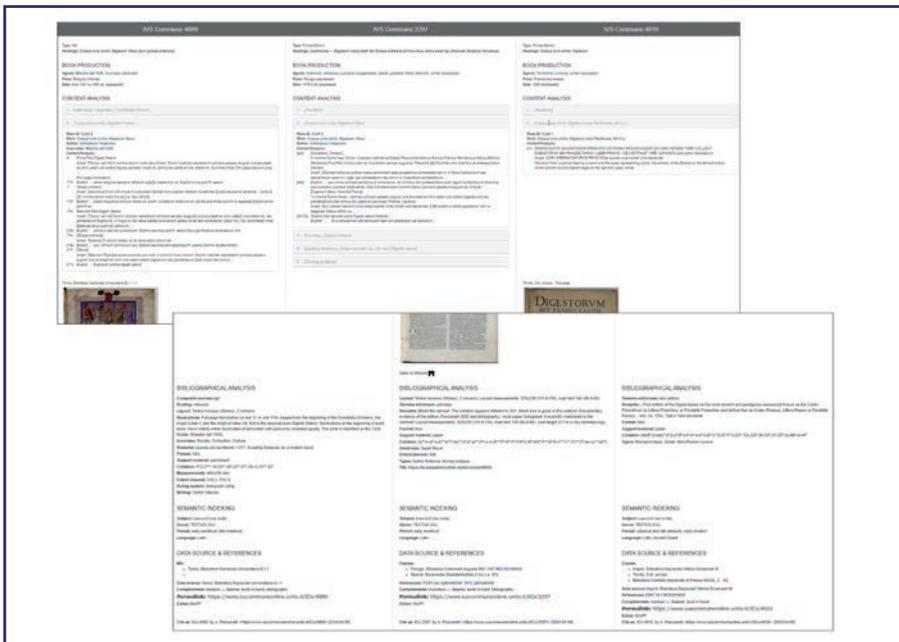


Figura 2. La visualizzazione simultanea delle descrizioni nel front-end

Questo risultato ha richiesto un lungo lavoro di progettazione della struttura del database e della presentazione dei dati. Nelle pagine che seguono se ne darà conto, illustrando a grandi linee la struttura generale del database, le caratteristiche peculiari dei record bibliografici, soffermandoci sulle sezioni principali che ne determinano la specificità e sulle relazioni con i record di autorità. Un primo paragrafo serve a dare conto di come il database è stato modellato e sviluppato, a partire

da una struttura esistente, quella del database TEXT-inc, su cui ICo ha modellato il suo nucleo descrittivo e da cui ha tratto un consistente pacchetto di dati.

2. Da TEXT-inc a IVS Commune *online*: disegno della struttura e importazione dati

TEXT-inc, come i lettori di questa rivista hanno avuto modo di leggere recentemente⁸, è la base dati dedicata ai testi stampati nelle edizioni del '400, nata sulla solida e ricca base costituita dalle oltre 5.500 descrizioni prodotte per il catalogo della Bodleian Library. Su quelle fondamenta il 15cBOOKTRADE ha continuato a costruire sviluppando una piattaforma destinata ad ospitare le descrizioni testuali delle restanti edizioni incunabile. È stato proprio lavorando in TEXT-inc, per produrvi la descrizione delle edizioni giuridiche del '400, che è nata l'idea di allargare e approfondire la ricerca in due direzioni: sull'asse del tempo, ampliando i limiti cronologici, così da includere descrizioni testuali tratte da manoscritti e da edizioni successive; in senso verticale, restringendo il campo al solo ambito giuridico.

L'analisi delle edizioni giuridiche del Quattrocento ha condotto, in breve tempo, ad alcune interessanti scoperte, mettendo bene in chiaro come le prime edizioni a stampa non costituiscono un insieme omogeneo e un mero momento di passaggio verso le più curate edizioni del Cinquecento⁹. Insieme ad alcune singole scoperte e ad uno studio complessivo sulla evoluzione delle edizioni giuridiche nei primi cinquant'anni di stampa¹⁰, il lavoro puntuale sui testi tratti da edizioni incunabile ha via via stimolato la necessità del confronto: con i testimoni manoscritti, per capire

⁸ La descrizione più aggiornata si trova in: Cristina Dondi — Matilde Malaspina, *L'ecosistema digitale del CERL per lo studio del libro antico a stampa: dal progetto 15cBOOKTRADE a oggi*. «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), n. 1, p. 134-156: 140-141.

⁹ Questa affermazione può sembrare persino superflua, e tuttavia l'idea che le prime edizioni a stampa, fatte salve alcune notissime eccezioni, non fossero altro che trasposizioni acritiche di testi reperiti nei primi manoscritti disponibili si è a lungo mantenuta e ha indotto talvolta o a ignorarle del tutto o ad effettuare selezioni funzionali alla costruzione di un campione che si vorrebbe rappresentativo del tutto. Si può osservare che questo atteggiamento si giustificava per il fatto che fino a non molti anni fa (diciamo oramai anche decenni) molte edizioni incunabile erano difficili da reperire, tante sono testimoniate in esemplari rarissimi e a volte unici. A maggior ragione, l'analisi testuale e la sua restituzione diventa un prezioso strumento che si affianca ai due grandi progetti bibliografici che forniscono le informazioni basilari sulle edizioni (ISTC e GW) e agli altri grandi progetti di studio (MEI, da un canto e il TW dall'altro, per citarne solo i principali) l'insieme dei quali può, a buon diritto, considerarsi un ecosistema.

¹⁰ I cui risultati si leggono in: Maria Alessandra Panzanelli Fratoni, *Printing the Law in the fifteenth century. With a focus on Corpus iuris civilis and the works of Bartolus de Saxoferrato*, in: *Printing R-evolution and Society. Fifty years that changed Europe (1450-1500)*, ed. by C. Dondi, Venezia: Edizioni Ca' Foscari, 2020, p. 67-198. Le singole scoperte hanno riguardato soprattutto la figura di Bartolo da Sassoferrato, la cui straordinaria fortuna ha significato anche la formazione di un ingente corpus di testi di dubbia attribuzione e quindi un interessante "problema editoriale", felice espressione coniata da Mario Ascheri, *Bartolo da Sassoferrato: introduzione a un giurista globale*, in: *Mundos medievales: espacios, sociedades y poder: homenaje al Profesor José Ángel García de Cortázar y Ruiz de Aguirre*, coord. Betriz Arízaga Bolumburu et al., Universidad de Cantabria: Editorial de la Universidad de Cantabria, 2012, p. 1029-1040: 1030.

quale lezione e quali arrangiamenti testuali fossero approdati in tipografia; con le edizioni più tarde, per verificare ad esempio eventuali legami o, al contrario, i cambiamenti di rotta¹¹.

Il punto di partenza, si diceva, è la base dati TEXT-inc, da cui IVS Commune ha derivato le logiche che governano il cuore della base dati, l'analisi dei testi, più tutto il contenuto ritenuto pertinente, cioè tutti i record relativi a edizioni giuridiche, che sono risultati circa un migliaio. Il passaggio dei dati è regolato da un accordo stipulato con Cristina Dondi; quanto, invece, alle responsabilità tecniche, sono tutte in capo a Luigi Tassarolo, che ha realizzato la nuova base dati e ne sta seguendo gli sviluppi¹². La selezione dei dati da importare, dalla base dati complessiva che ne conteneva oltre 8.000, è stata effettuata con una sorta di triangolazione, passando cioè per l'intero insieme di edizioni giuridiche del '400, estratto dalla base dati ISTC. Si è cercato in questo modo di ottimizzare lo sforzo richiesto dalla importazione dei dati, tenuto conto anche del fatto che la nuova base dati è realizzata con un diverso sistema. Si è quindi cercato di allestire un insieme iniziale di record che fosse il più ampio possibile e già organizzato secondo la nuova struttura.

Il punto merita qualche riga di commento: si disponeva della estrazione dei dati presenti in TEXT-inc, da cui si dovevano selezionare i soli record pertinenti, ovvero tutte e solo le edizioni che ricadono in ambito giuridico. D'altro canto, si disponeva di una estrazione complessiva di dati da ISTC, tutti già identificati per ambito disciplinare¹³: circa 4.800 record.

Per fare sì che in IVS Commune confluissero i record di tutta la produzione a stampa del Quattrocento con la maggiore quantità possibile di dati, si è deciso di preparare un file master con un foglio di calcolo che fungesse, al contempo, da disegno della nuova struttura e contenesse tutti i dati da importare nelle forme desiderate. L'identificativo ISTC (presente in entrambe le basi di dati) è servito anche come filtro per recuperare, da TEXT-inc, i soli record di interesse, cir-

¹¹ Gli sviluppi della editoria giuridica nel Cinquecento si preannunciano sul finire del secolo precedente con l'attività di tipografi come i Torti, ad esempio, il cui ruolo andò ben oltre la produzione di un ricco catalogo specializzato. A loro si deve verosimilmente la "iniziazione" dei Giunta alla stampa di testi di diritto, ambito nel quale avrebbero imperato nel corso del secolo.

¹² A Tassarolo si deve la scelta di un diverso sistema di gestione della base dati (DBMS), passando da PostgreSQL, utilizzato per costruire TEXT-inc, mentre IVS Commune *online* è sviluppato in MySQL (per la precisione MariaDB, che è open source). La scelta, operata da Tassarolo, è dettata da ragioni di carattere pratico e tiene conto delle dimensioni contenute della base dati, d'altra parte felicissimo si rivela il dialogo con lo sviluppatore che è ben noto alla comunità dei "digital humanists" italiani (a lui si devono i due db del progetto EMoBookTrade; la realizzazione dell'illuminated Dante Project e, sulla stessa base di ICo, anche il più recente AGAPE).

¹³ Una prima indicizzazione semantica di tutto quanto resta delle edizioni incunabile (circa 30.000) è tra i risultati del progetto 15cBOOKTRADE, che ha visto impegnato per molti mesi il gruppo di lavoro; il tema è ripreso più avanti nel paragrafo dedicato. Da quella classificazione emergono 4.615 edizioni di testi giuridici a cui sono state aggiunte circa 200 edizioni di testi identificati come giuridici in via secondaria, come le orazioni di Cicerone, che erano state etichettate come "letteratura (classica)" ma che hanno una indubbia origine giuridica.

ca poco più di mille, da cui è stato importato il nucleo, il cuore testuale della descrizione.

È stato invece importato integralmente l'indice dei nomi degli autori (principali e secondari, un deposito molto ricco, frutto delle analisi di circa 8.000 edizioni, e della relativa identificazione o comunque indicizzazione delle responsabilità intellettuali). Quanto alla nuova struttura, il disegno è stato impostato in direzione di: un forte potenziamento di alcuni aspetti della descrizione, nella fattispecie una intera sezione destinata ai dati relativi alla descrizione materiale del codice (manoscritto e a stampa); nella previsione di una serie di elementi di valutazione critica, relativi ad esempio alla certezza dell'informazione (per le date, topiche e croniche; per le attribuzioni) o anche alla ipotetica posizione dell'edizione nello stemma. È stato previsto un forte potenziamento degli indici, con l'aggiunta di un titolo uniforme per le opere e l'allestimento di un relativo record di autorità ma anche in relazione ai nomi di stampatori, editori e officine tipografiche. Nel master si è poi intervenuto per corredare i dati già presenti con elementi di valore appunto previsti nella nuova base dati: i termini di indicizzazione semantica elaborati appositamente (sui quali si veda il par. 6), la toponomastica, con l'aggiunta degli identificativi univoci per i luoghi e paesi; sono pure stati lavorati in massa gli attributi relativi a certezza o incertezza del dato.

Creata la nuova struttura, la base dati è stata così popolata con un caricamento massivo di circa 4.800 record, in parte completi di descrizioni testuali, in parte pronti per riceverle. Nella struttura così allestita si è continuato a lavorare, arricchendo la sezione dedicata alla descrizione bibliografica e affinando quella relativa ai testi, con modalità e risultati descritti nei paragrafi che seguono.

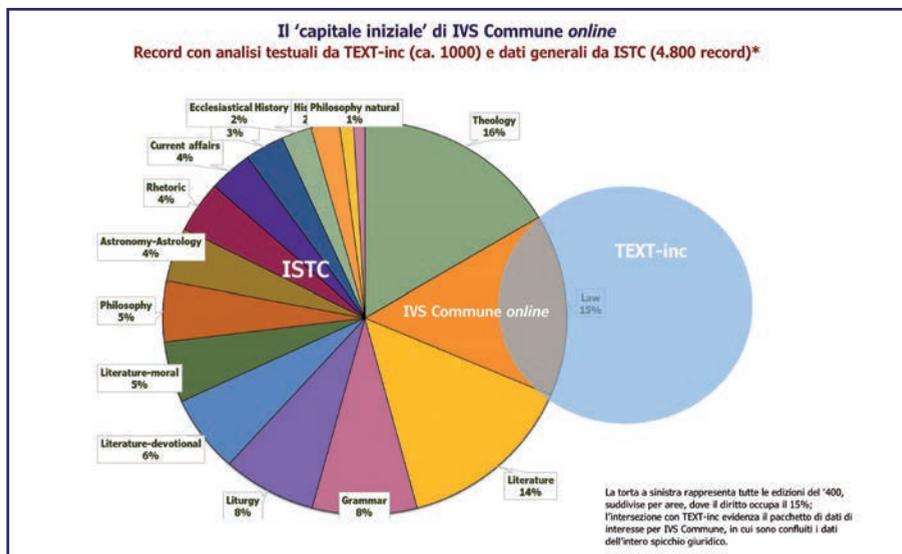


Figura 3. Il sottoinsieme bibliografico del primo popolamento massivo della base dati

3. Opere, testi e immagini (tra i testi)

Il cuore della base dati è la sezione dedicata ai testi, identificata nel back-end come “Content Analysis”. La sezione si presenta come una stringa vuota destinata ad ospitare i testi e i paratesti di una o più opere, tutte quelle contenute nel volume che si sta descrivendo. Selezionando la stringa vuota, si apre una finestra già predisposta per ospitare le trascrizioni di ogni unità rilevante, ovvero *incipit* ed *explicit* di ogni unità testuale, alternate alle trascrizioni integrali di titoli, colophon, dichiarazioni di privilegio e altri apparati paratestuali. La trascrizione di *incipit* ed *explicit* è il minimo richiesto, non un limite: l’editor ha la libertà di produrre trascrizioni integrali di tutte le sezioni che ritenesse rilevanti. Un campo ospita ciascuno di questi elementi e l’editor può aggiungere tanti campi quanti ritiene necessari, esplicitando ogni volta la posizione nel volume con riferimento alla carta (locus). Insieme alle trascrizioni, in questa sezione si inseriscono le voci d’indice, per le responsabilità intellettuali aggiuntive rispetto a quella dell’opera principale; per questa, infatti, la responsabilità autoriale è dichiarata nel record di autorità dell’opera, verso il quale si costruisce la relazione attraverso una sottosezione dedicata: Uniform Heading (Work). Ogni opera viene identificata tramite un titolo uniforme, che viene recuperato direttamente da una scheda di autorità in cui sono riportati anche i titoli varianti attestati in letteratura, i legami agli autori, la bibliografia ed eventuali altre informazioni¹⁴.

Completata la trascrizione di ogni elemento relativo ad ogni opera si fa il percorso inverso fino a tornare alla prima schermata della sezione, dove è possibile aggiungere altri testi, se si sono rinvenute altre opere. In questa fase è anche possibile agire sulla sequenza dei blocchi di testo, per spostarli in posizioni precedenti o successive, se ciò è reso necessario da eventuali errori o semplicemente perché si lavora di fronte ad un volume miscelaneo (un manoscritto unitario o una miscellanea editoriale) che non si intende descrivere ordinatamente dall’inizio alla fine.

In questa sezione, come probabilmente si è già intuito, un’operazione delicata e non sempre facile è l’identificazione dell’opera: non tutte le opere, infatti, sono attestate nei repertori e identificabili con titoli uniformi. Non sempre, inoltre, le opere sono chiaramente identificate e le relative responsabilità acclamate.

Analizzate in questo modo, inoltre, si riconoscono come opere a sé stanti anche quegli elementi degli apparati para-testuali che hanno carattere di autonomia e proprie responsabilità autoriali, quali lettere prefatorie, biografie degli autori e simili.

Un approccio così analitico mira a restituire la complessa struttura di testi e paratesti in cui un’opera si presenta nelle sue diverse versioni, e mira a farlo in modo che le descrizioni siano funzionali allo studio della trasmissione del testo anche

¹⁴ L’applicazione di un record di autorità alle opere e l’aver distinto i due momenti, descrizione dei testi e indicizzazione delle opere, è l’elemento che più differenzia il modello proposto da ICo rispetto a TEXT-inc, dove i titoli delle opere sono identificati da un’etichetta dedicata (work appunto) e trascritti allo stesso livello dei singoli testi.

per impatto visivo, per il modo in cui si presentano. Nella visualizzazione del record, le singole opere sono riunite in pannelli, aprendo i quali si possono leggere le trascrizioni dei relativi testi.

Nell'analisi dei testi rientra anche la registrazione della presenza di apparati figurativi, miniature e vignette alla descrizione delle quali è dedicato ampio spazio nella sezione Bibliographical Analysis, ma che, essendo anch'esse veicolo di significati, sono richiamate nel corso della descrizione testuale. Al raccordo tra apparati figurativi e contenuto testuale, ICo pone molta attenzione, cercando un confronto costante con chi si occupa specificamente di questi temi¹⁵. Gli apparati iconografici portatori di significato (miniature, iniziali figurate o istoriate ecc.) si trovano così segnalati nei loci in cui compaiono e in associazione ai testi, così che i loro significati ne risultino esaltati e valorizzati. Naturalmente le immagini sono frutto di tecniche specifiche e si analizzano anche per gli aspetti materiali.

In modo analogo, ed è ben noto, i testi comunicano anche (forse anzitutto) per il modo stesso con cui sono disposti sulla pagina, ovvero grande importanza ha la *mise-en-page* e peculiarissima è la *mise-en-page* nel libro giuridico, che nell'età d'oro dello *ius commune* si fissò nelle forme di un *textus inclusus*, con il testo autorevole delle leggi (romane e di quelle canoniche, che sulle prime esemplarono l'arrangiamento dei testi) collocate al centro della pagina, e incorniciate dalla glossa ordinaria. Questa *mise-en-page*, che il diritto condivise con altre discipline insegnate a livello universitario (la teologia in particolare) era appunto il risultato delle attività di esegesi e insegnamento poste in essere nelle università (altra grande invenzione istituzionale del basso medioevo) che si rispecchiò nell'allestimento di un libro particolare, che sarebbe cambiato di nuovo quando un diverso approccio al testo cominciò a farsi strada. Avvenne nel passaggio all'età moderna, ed avvenne per effetto di un approccio diverso, e critico, al testo delle *auctoritates*; avvenne infatti quando anche in ambito giuridico si fece sentire il vento dello sguardo filologico degli umanisti, dando avvio a quel movimento che è noto come umanesimo giuridico. Il rilevamento della *mise-en-page* è quindi essenziale ai fini che ICo si propone, è un elemento spia della tipologia di libro cui ci si trova di fronte, anche del suo grado di formalizzazione e di ufficialità. Componente propria del sistema di comunicazione, la *mise-en-page* è però anche un elemento dell'allestimento materiale dei supporti. La sua è quindi inserita nella sezione dedicata al rilevamento e registrazione degli aspetti materiali, come sopra anticipato, sezione Bibliographical Analysis.

¹⁵ In particolare, esiste una collaborazione con Maria Alessandra Bilotta e la sua "oficina de investigação" IUS Illuminatum, progetto dedicato allo studio dei manoscritti giuridici miniati (<<https://iusilluminata.fcsh.unl.pt/>>). Il dialogo con Bilotta si esplica ora anche con un co-tutorato nel dottorato di ricerca che Camilla Marangoni conduce nell'ambito del Dottorato Nazionale in Heritage Science (curriculum Archivi e Biblioteche); parte della ricerca, che affronta nel suo insieme gli incunabili miniati nelle collezioni torinesi, è dedicata allo sviluppo di un linguaggio iconografico comune per la descrizione degli apparati iconografici nel libro giuridico.

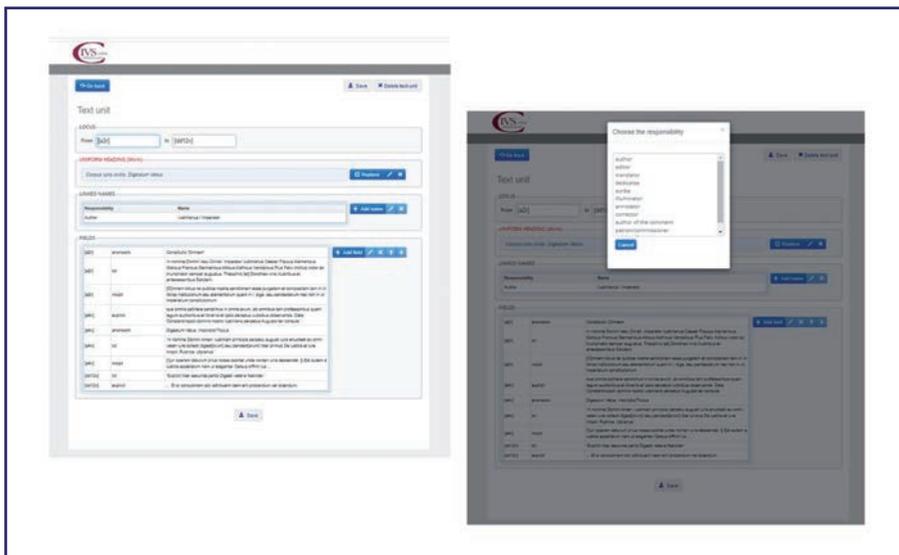


Figura 4. “Content Analysis”: opere, testi e legami agli indici (vista back-end)

4. Mise-en-page, collazione, supporti, misure: la struttura del codice (la sezione Bibliographical Analysis)

Analisi bibliografica vuol dire analisi del libro per tutti i suoi aspetti materiali, dunque analisi del codice; ai lettori di questa rivista non serve ricordare che nella grande rivoluzione che è stata l'introduzione della stampa, una cosa non è cambiata ed è appunto la struttura del libro. Lo si apprezza bene quando ci si trova a studiare insieme libri manoscritti e a stampa, al fine di rilevare i momenti di continuità e quelli di discontinuità; tali momenti interessano i testi veicolati ma si rilevano bene, e anzitutto, nella struttura del libro stesso. La sezione registra quindi in primo luogo formato del libro, supporti e misure; il formato è rilevato anche per il foglio tipografico, di cui un numero crescente di studi riconosce l'importanza; in particolare per il libro giuridico si rileva, per il primo periodo, l'uso frequente di un formato mediano tra i due maggiori, imperiale e reale, Super-royal l'ha denominata Paul Needham¹⁶.

Di seguito si registra la collazione, rispetto alla quale si è invitati a dire se è regolare o no (selezionando una opzione); regolarità e irregolarità della collazione sono infatti indizi, ad esempio di una edizione nuova e indipendente. La collazione – lo hanno messo in evidenza studi condotti sulla stampa del '400 – può anche svelare l'esistenza di un programma editoriale, quando ad esempio si rinviene una sequenza di simboli utilizzati dal tipografo per le diverse parti di un *corpus*. Com'è noto, nel corso del medioevo, le quattro compilazioni giustiniane (Digesto,

¹⁶ Paul Needham, *Aldus Manutius's Paper Stocks: The Evidence from Two Uncut Books*, «The Princeton University Library Chronicle» 55 (1994), n. 2, p. 287–307: 292–293.

Codice, Istituzioni e Novelle) furono divise e riarrangiate in sezioni e parti, funzionali anche al loro insegnamento¹⁷. Su tali divisioni (che prevedevano altresì un'ulteriore suddivisione in prima e seconda parte di ciascuna delle sezioni del Digesto e del Codice) si sarebbe appuntata poi, con la Scuola dei Commentatori, l'opera esegetica dei giuristi, formando *Lecturae* e poi *Commentarii* al *Digestum Vetus*, *Infortiatum*, *Novum* (o anche solo alla prima o alla seconda parte di ognuno) al *Codex* e così via.

La divisione in parti era riflessa nelle unità codicologiche e poi, con la stampa, in unità bibliografiche, talché nei primi decenni si hanno le diverse parti stampate in diverse edizioni. Interessante, a questo proposito, notare come le *editiones principes* di ogni sezione del *corpus* giustiniano venissero realizzate in luoghi diversi, e come poi con il tempo alcuni stampatori/editori si attrezzassero per stampare i *corpora* al completo, come pure i *commentarii* dei principali giuristi a tutte le parti. Anche in questi casi, ogni volume veniva solitamente concepito come edizione a sé stante, ed è allora che la segnatura dei fascicoli può diventare rivelatrice di un'operazione editoriale complessiva, laddove essa si presenta in continuità tra i volumi. Allo stesso modo si può ricostruire la serie dei legami esistenti tra edizioni diverse parte della stessa opera, a questo scopo è predisposto un campo per evidenziare il legame (campo Related editions).

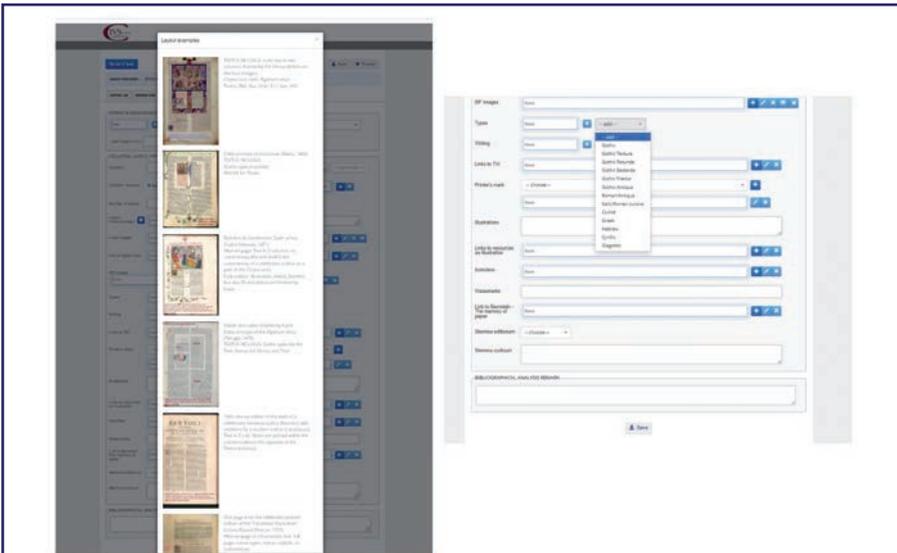


Figura 5. "Bibliographical Analysis" (back-end): mise-en-page e parti di descrizione

¹⁷ I 50 libri del Digesto furono divisi in tre: *Digestum Vetus* (1-24.2), *Infortiatum* (24.3-38) e *Novum* (39-50); il Codice (12 libri) in due: i libri 1-9 formavano il *Codex*, i libri 10-12, denominati *Tres Libri*, erano solitamente legati con i quattro libri delle *Institutiones* (il testo base per la formazione degli studenti) e con le *Novellae Constitutiones*, a formare il *Volumen parvum*.

Numero di carte, di righe e sistema di rigatura (quest'ultima per i manoscritti) corredano il campo dedicato alla mise-en-page (nella base dati Layout), alla quale come si diceva si riconosce grande importanza; onde facilitare il lavoro dell'editor, il campo è a scelta multipla di elementi precompilati (Fig. 5).

La descrizione, naturalmente, non è perfettamente efficace nel restituire l'aspetto della pagina di cui però si possono fornire esemplificazioni tramite alcuni campi dedicati alle riproduzioni digitali. La sottosezione dedicata alle immagini offre tre distinte possibilità, in ordine di preferenza: legami verso oggetti digitali riprodotti secondo lo standard IIIF, ciò che consente di legare al record l'immagine di qualche pagina significativa e di sfogliare l'intero volume agganciandosi al visualizzatore.

Quando la riproduzione in IIIF non esiste, ma esiste una riproduzione comunque integrale, si opta per un link verso quest'ultima. In entrambi i casi vi è la possibilità di agganciare anche un'immagine locale, realizzata dall'editor, qualora non esistesse una riproduzione presso l'istituto di conservazione oppure quando si volesse comunque produrre l'immagine di un altro esemplare. Il legame verso digitalizzazione in IIIF è la scelta preferita perché questo consente di usufruire di immagini di grande qualità e leggibilità senza costi aggiuntivi per il progetto, senza gravare sul server su cui insiste il database, e sfruttando al meglio (quindi anche valorizzando) le operazioni di digitalizzazione massiva poste in essere da alcuni dei più importanti istituti di conservazione, come la Biblioteca Apostolica Vaticana, la Bayerische Staatsbibliothek, le Bodleian Libraries e un numero via via crescente di biblioteche. Utile complemento alla descrizione della mise-en-page, le immagini servono bene a descrivere, oltreché esemplificare, i sistemi grafici e tipografici, per entrambi i quali l'editor sceglie tra gli elementi di una lista preordinata, che si può incrementare secondo le esigenze, anche seguendo i suggerimenti di integrazione o correzione che gli specialisti paleografi faranno pervenire; al momento la lista contiene le definizioni più comunemente rilevate in relazione al testo giuridico (gotica, gotica rotunda, *littera bononiensis* ecc.). Completa si può invece considerare la lista delle definizioni dei tipi, che si appoggia sul *Typenrepertorium der Wiegendrucke* (TW), le cui ricchissime schede sono puntualmente richiamate con un campo dedicato e ripetibile¹⁸.

Quasi riproducendo l'evoluzione tra l'attività delle prime officine, studiata attraverso l'esame dei tipi, e quella di stampatori/editori che adottarono un proprio emblema, segue una sottosezione riservata alla descrizione delle marche tipografiche, con un campo ripetibile per creare legami verso i repertori online, a partire dai record d'autorità dedicati alle marche in Edit16; naturalmente il riferimento ai repertori cartacei sarà incluso nel campo bibliografia.

¹⁸ Nato come sviluppo digitale del celebre repertorio creato da Conrad Haebler, TW sembra sviluppare splendidamente le possibilità offerte dallo strumento digitale facendo nascere, dalla panoramica dei tipi, una ricostruzione delle relazioni tra le persone attive nelle officine (<<https://tw.staatsbibliothek-berlin.de/>>).

Appena dopo la sezione dedicata alle marche c'è quella delle illustrazioni, ovvero lo spazio per descrivere apparati iconografici e decorativi, nei termini di tecniche, tipologie, misure, che integra così le interpretazioni fornite degli stessi apparati nell'analisi del contenuto. Alle illustrazioni è inoltre applicato il sistema di classificazione ICONCLASS¹⁹, per codificare le descrizioni prodotte dall'editor mitigandone la soggettività e potenziando lo studio comparato, sia in direzione sincronica, tra testimoni prodotti nello stesso periodo, che diacronica, per seguire anche le evoluzioni degli apparati iconografici nel corso del tempo. L'adozione di ICONCLASS è fatta anche per analogia con le modalità di descrizione delle illustrazioni sviluppata da 15Illustration, il sistema digitale creato per studiare in modo sistematico e oggettivo uso e riuso di matrici e altri materiali utilizzati in tipografia per gli apparati decorativi. È altresì prevista la possibilità di creare legami diretti, tanti quanti ne servono, con le descrizioni presenti nel 15cIllustration con un campo che rinvia appunto a progetti dedicati²⁰.

L'applicazione di ICONCLASS a una particolare tipologia di testi e di libri crea l'occasione per lavorare alla elaborazione di un linguaggio dedicato, potenziando il dialogo interdisciplinare, in questo caso tra storia del libro, storia del diritto e storia della miniatura; da tale dialogo è emersa l'occasione di uno sviluppo di ICONCLASS in direzione di una maggiore e migliore rappresentatività della sezione giuridica²¹. Due campi sono infine predisposti per accogliere le descrizioni delle filigrane e corredarle di legami verso quelle che si fossero trovate nel catalogo del portale Bernstein – The memory of paper²².

A chiudere la lunga sezione dedicata all'analisi bibliografica ci sono due campi progettati per informare circa la posizione che il volume occupa (o si ipotizza possa occupare) nello stemma – *codicum* o *editionum* a seconda che sia tratti di un manoscritto o un'edizione a stampa. Naturalmente non si pretende che dall'esame di un testimone manoscritto possa derivare la conoscenza della sua posizione nello stemma, per i manoscritti è disposto un campo a testo libero in cui registrare ogni dato utile (inclusa ovviamente la posizione assegnata dalla letteratura). Più semplice appare invece collocare l'esemplare a stampa, per il quale sono previste quattro semplici opzioni: *princeps*, *new edition*, *dependent edition*, *first edition*;

¹⁹ Una bibliografia aggiornata sul sistema di classificazione si trova nelle pagine introduttive del sito web: <<https://iconclass.org/help/abouta>>.

²⁰ Prevedendo l'espansione di 15cIllustration verso i secoli successivi; in poche parole, 15cIllustration è uno strumento digitale finalizzato allo studio sistematico di vignette e altri elementi iconografici, che sfrutta un software di image-matching per analizzare con precisione analogie e differenze tra i singoli elementi decorativi preventivamente catalogati in modo puntuale. Una sintesi e riferimenti bibliografici si trovano in: C. Dondi — M. Malaspina, *L'ecosistema digitale del CERL cit.*, p. 141-142.

²¹ Camilla Marangoni, già citata per la ricerca dottorale, sta ora svolgendo il tirocinio all'estero previsto dal Dottorato Nazionale presso lo Huygens Instituut di Amsterdam in stretto dialogo con Hans Brandhorst, responsabile scientifico di ICONCLASS.

²² Portale dedicato allo studio della carta che contiene un ricchissimo catalogo digitale di filigrane: <www.memoryofpaper.eu>.

princeps e *first edition* non dovrebbero avere bisogno di spiegazione, mentre *new edition* e *dependent edition* servono a distinguere la nuova edizione da una edizione pure nuova ma esemplata su una precedente.

Con l'applicazione di queste etichette su tutte le edizioni registrate ci sembra si possa creare una base per provare a districare il complicato intreccio delle edizioni di testi giuridici che cominciarono ad affollarsi già nel Quattrocento e che nel corso del secolo successivo formarono un intrico quasi impossibile da riordinare. Naturalmente tali valori possono appoggiarsi ai tanti studi che esistono in merito, sia relativamente a singoli autori sia a ricostruzioni complessive; una ricognizione della letteratura scientifica è anzi il punto di partenza di ogni nuova descrizione. È inoltre probabile che l'editor debba produrre elementi a supporto dei giudizi espressi, ciò che può fare in un campo per eventuali note relative all'analisi bibliografica (Remark). L'attendibilità delle informazioni fornite si può giudicare, anche lato pubblico, tenendo insieme queste eventuali note con quanto dichiarato nella sezione *References*, dove si esplicitano le fonti delle informazioni (libro in mano/letteratura/cataloghi) e si dichiara il livello (minimo, medio o massimo) della descrizione.

5. Responsabili del testo e responsabili del libro

Si dedica un paragrafo unico a descrivere le modalità con cui in ICo sono registrate le diverse responsabilità, quelle relative ai testi e quelle legate al manufatto, che sono indicizzate da sezioni diverse della scheda ma la cui identificazione ha comunque un inizio (o un momento di verifica) nell'analisi e trascrizione dei testi.

L'indicizzazione delle responsabilità intellettuali, relative cioè ad opere e testi, avviene dalla sezione dedicata all'analisi del contenuto, come si è già accennato; spendiamo qui ancora qualche riga per dire come si presenta il record d'autorità e per gli attributi associati alle responsabilità, e che mirano ad ottimizzare il lavoro di analisi e identificazione.

Circa il record di autorità, questo si presenta sostanzialmente identico a quello della base dati TEXT-inc, che si caratterizza per una struttura articolata e ricca di informazioni: oltre alle informazioni consuete (estremi biografici, genere, varianti onomastiche) sono fornite informazioni sullo status, la professione, la localizzazione e il periodo di attività. In ICo sono stati aggiunti i riferimenti agli identificativi ISNI e VIAF²³, per potenziare la precisione nella identificazione e accrescere così le occasioni di scambio di dati con altri sistemi. Molto diversa, rispetto a TEXT-inc, è la modalità di creazione delle relazioni con il record bibliografico. Il legame all'authority, in ICo, è costruito dalla sezione Content Analysis, muovendosi tra i due livelli: si entra in un livello superiore nel momento in cui si registra il titolo uniforme e lì si crea la relazione all'autore; le relazioni verso le altre re-

²³ International Standard Name Identifier (<<https://isni.org/>>) e Virtual International Authority File (<<https://viaf.org/>>), entrambi ben noti a chi utilizza strumenti bibliografici online.

sponsabilità – in linguaggio FRBR potremmo dire le responsabilità individuate con riferimento all’espressione e, in parte, alla manifestazione – si effettua nella fase di trascrizione dei testi.

Nel creare il legame, si deve (si può) specificare se la responsabilità è certa (acclamata e pacifica), se è dubbia, se si tratta di una pseudo-autorialità o, infine, se il soggetto citato è autore di riferimento. Esempi di questa ultima occorrenza sono i repertori di opere di autori celebri (come, ad esempio, il *Repertorium Bartoli* di Antonio Mincucci) o anche le *Additiones* che autori della primissima modernità fecero ai grandi giuristi dell’età mediana; di nuovo un esempio ci è offerto da Bartolo sulla cui produzione esegetica scrisse numerose *Additiones* Alessandro Tartagni (m. 1477). La possibilità di esplicitare il grado di certezza dell’attribuzione autoriale (naturalmente basata o comunque suffragata dalla letteratura) consente di superare alcune contraddizioni che è capitato di rilevare tra i cataloghi, quando a fronte di una tradizione obiettivamente complessa, le opere dubbie si trovano talvolta attribuite, talaltra legate a uno pseudo, talvolta ancora attribuite con una nota di correzione, con risultati tra loro discrepanti²⁴.

A livello della specifica edizione che si sta descrivendo si avrà una serie di altre responsabilità che vengono indicizzate creando un legame ai relativi record di autorità per ciascuno dei quali si esplicita il ruolo svolto: editore scientifico, commentatore, traduttore, come pure dedicatario, committente, pubblico ufficiale (per i casi ad esempio di dichiarazioni formali, come sono i privilegi, le autorizzazioni alla pubblicazione ecc.).

Naturalmente, dalle descrizioni dei colophon emergono stampatori ed editori, i cui nomi però vengono registrati in un’altra sezione, denominata Book Production, dove il focus è appunto sui dati relativi alla manifattura del libro. In questa sezione si ripetono, in un campo unico, i dati così come si trovano nei cataloghi generali (con i quali si vuole rimanere allineati), questi vengono poi corredati da informazioni puntuali dei nomi degli individui (di copisti, stampatori, editori), e insieme delle denominazioni delle officine tipografiche e delle case editrici: per il Quattrocento ci si appoggia ancora una volta al rilevamento condotto a Berlino dall’équipe di GW; per i secoli successivi di nuovo ad Edit16, per l’Italia, e al Thesaurus del CERL per uno sguardo europeo. Oltre ai nomi si indicizzano, naturalmente, i luoghi di stampa, identificati anche con l’identificativo geonames e i paesi (secondo la geografia attuale); infine le date, che si possono registrare con precisione al giorno, se presente, oppure circoscritte in un intervallo, se incerte. Per ogni evenienza esiste un’opzione da selezionare.

²⁴ Ancora Bartolo da Sassoferrato offre esempi utili, con riferimento a un colorito trattato, connotato da una complessa e duplice tradizione, che gli è talvolta attribuito e rispetto al quale la critica non ha prodotto un giudizio definitivo. Ci si riferisce qui al *Processus Satanae*, che alcuni dei principali repertori hanno trattato in modo disomogeneo. Il tema è illustrato in: M.A. Panzanelli Fratoni, *Printing the Law*, cit., p. 193-194.

Il grado di certezza viene inoltre esplicitato per ciascuno degli elementi registrati: nomi, luoghi, date. Si deve selezionare una delle opzioni disponibili: dato espresso, parzialmente espresso, inferito, attribuito, corretto. Questo consente di chiarire meglio alcune informazioni (come attribuzioni non pacifiche o controverse) e di utilizzare i livelli di incertezza che altrimenti non si riescono a recuperare (benché implicitamente dichiarati dall'uso delle parentesi quadre). In quest'area è presente, inoltre, un campo per proporre datazioni alternative a quelle ufficialmente accettate dai principali repertori (alternative data): non è infrequente, infatti, che studi puntuali sulla produzione di un luogo o di uno stampatore/editore riescano a precisare, integrare, correggere dati acquisiti e apparentemente consolidati. Sempre in questa sezione si trova il campo dedicato al *secundo folio* (eredità di TEXT-inc che volentieri si mantiene), ovvero le prime parole della prima riga del secondo foglio redatto (elemento identificativo dei singoli volumi usato negli inventari medievali, funzionale alle identificazioni rispetto alle collezioni antiche).

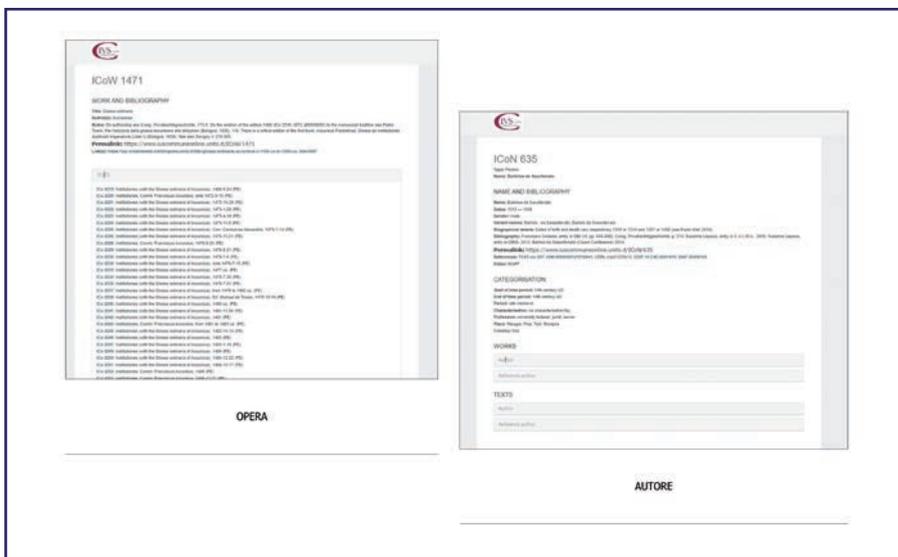


Figura 6. I record di autorità per opere e autori (vista front-end)

6. Indicizzazione semantica

Un ultimo importante passaggio nell'analisi dei testi consiste nell'indicizzarli con riferimento al contenuto in relazione ad alcune categorie di ambito, non soggetti in senso stretto, piuttosto aree disciplinari, elaborate sulla base di repertori specializzati. L'operazione in sé riprende in mano, per specializzarla, quella effettuata sull'intera produzione del '400, condotta nell'ambito del progetto 15cBOOKTRADE, quando si è trattato di analizzare le circa 30.000 edizioni incunabile che si conservano oggi. Queste furono divise in categorie identificate con

descrittori elaborati tenendo conto delle esigenze degli studiosi di oggi e al tempo dei sistemi di sapere e della terminologia dell'epoca. Si riuscì così ad ottenere una significativa visione d'insieme della produzione a stampa del '400, raggruppata in categorie piuttosto ampie ma restituita nella sua ricchezza dall'applicazione di una serie di parole chiave finalizzate a mettere in chiaro aspetti come l'ambito scolastico (ad esempio le grammatiche o i testi per lo studio universitario); la presenza di testo poetico; la tipologia per testi di consultazione (i dizionari ad esempio). In quel contesto per l'area giuridica furono individuati tre sottogruppi: le due categorie classiche di diritto civile e diritto canonico (Law-civil e Law-canon) più una terza categoria "pratico" (Law-practical), con ciò intendendo testi legati alla pratica notarile, all'amministrazione, opere appunto di carattere pratico piuttosto che dottrinario.

Concentrandosi sulla sola produzione giuridica l'Co può e vuole sviluppare un sistema più ricco di distinguo e sfumature, capace di restituire la complessità di una produzione testuale che insieme ai grossi volumi della dottrina o delle leggi, include anche brevi testi normativi, atti di amministrazione, raccolte di consuetudini ecc. I descrittori sono stati dunque arricchiti per aggiungere, alle definizioni di diritto civile e diritto canonico, quella di *utrumque ius*; per consentire una distinzione, nel gruppo dei testi a carattere pratico, tra le opere indirizzate ai notai, come i formulari, e le procedure; per consentire di descrivere meglio i testi dei cosiddetti *iura propria*, cioè i diritti particolari espressione di singole comunità, territoriali e non: consuetudini, statuti di città ma anche statuti di comunità più ristrette, come le corporazioni, i collegi ecc. Infine per prevedere i casi delle emanazioni normative di gruppi nazionali così come quelli degli stati di età moderna.

Attualmente i descrittori in uso sono dodici, espressi sia in inglese, sia in latino (per la maggiore perspicuità dei termini latini in questo ambito) e includono: *ius canonicum*; *ius civile*; *utrumque ius*; *ius practicum*; *ius processuale*; *ius proprium (civitas / castrum / villa)*; *ius proprium (collegium / fraternitas / nationes / societas / universitas)*; *ius proprium (consuetudines)*; *ius proprium (natio)*; *l'ars notariae* e infine anche un "*rhetorica iuridica*" per quei testi che, nati in ambito giuridico, sono poi riproposti come modelli di retorica appunto (si pensi alle orazioni ciceroniane). Ai descrittori semantici, che non sono esclusivi (se ne possono applicare tanti quanto è necessario) si aggiungono termini che definiscono il genere, la tipologia di testo (*genres*), per distinguere testi delle leggi vere e proprie (*textus*) dalla dottrina, i *commentarii* dai *tractatus*, e una serie di altri termini, tratti dalle edizioni stesse e dalle bibliografie specializzate dell'epoca. Qui vi sono anche termini che indicano le compilazioni di testi come antologie e miscellanee, indicazione molto utile a tracciare i paesaggi letterari di lungo periodo; le compilazioni, in particolare, possono rivelarsi ottime spie di continuità o discontinuità tra una tradizione e l'altra.

Associato ai periodi di attività di autori primari e secondari legati a ciascuna edi-

zione, l'applicazione dei descrittori semantici e di genere serve di supporto alla ricostruzione delle trasformazioni del paesaggio giuridico europeo per come fu testimoniato dai testi: dai corposi volumi della dottrina, ai fogli volanti contenenti singoli atti normativi. Si tratta evidentemente di una sfida, che può provocare reazioni da parte di chi conosce la disciplina dal suo interno. È in verità quanto si spera: stimolare il dialogo con gli storici del diritto per sperabilmente instaurare forme di cooperazione.

7. Cooperazione, interoperabilità, networking

La cooperazione è essenziale perché un progetto come quello qui illustrato possa ambire a risultati significativi; più in generale, potremmo dire che la cooperazione è essenziale per progetti di ambito bibliografico, che è per sua stessa natura vasto, complesso e difficilmente confinabile entro spazi controllabili da una sola persona. Non dovrebbe esserci bisogno di aggiungere considerazioni a supporto, ci sono già diversi progetti il cui successo chiaramente dipende da un buon coordinamento delle tante parti chiamate a lavorare insieme. Certo, un buon coordinamento è assai più facile a dirsi che a farsi e i risultati dipendono anche dalla soddisfazione di tutte le parti, e non sempre risultano a tutti immediatamente evidenti. Ciò rende tanto più ammirevole il successo di progetti di lungo periodo come Edit16, o SBN²⁵. Una sfida ancora diversa è quella affrontata da progetti che mirano a far dialogare i professionisti, curatori e bibliotecari, con chi sui libri basa la propria ricerca scientifica. È il caso di MEI (Material Evidence in Incunabula), i cui risultati sono a tutti evidenti: centinaia di editor che lavorano nella stessa base dati, col risultato di valorizzare le raccolte più piccole portandole su un palcoscenico globale insieme alle collezioni più grandi²⁶.

La cooperazione che IVS Comune *online* mira a promuovere si radica sull'esperienza maturata nel collaborare ad alcuni dei progetti sopra citati, e sulla osservazione che un primo grande risultato si ottiene già facendo convergere, in un unico spazio, ricerche condotte in modo indipendente. Si crea così un punto di partenza per una fase successiva, in cui più persone lavorano contemporaneamente nella raccolta dei dati. Lavorando sui testi giuridici di età tardomedievale è stato naturale confrontarsi con la produzione manoscritta, ambito nel quale esiste un lavoro davvero pionieristico di raccolta dei dati che ha sfruttato precocemente le tecnologie

²⁵ Sul successo di SBN: Simonetta Buttò, *Trent'anni di SBN*, «AIB studi» 56 (2016), n. 2, p. 181-183 e l'intero numero di «DigItalia» del 2017.

²⁶ In MEI lavorano insieme professionisti del libro che hanno alle spalle una lunghissima carriera e studenti alle prime armi, per i quali il lavoro nella base dati è un grande esercizio e l'inizio di un percorso. Va sottolineato che a ricordare che le biblioteche sono un laboratorio di ricerca sono tutte le attività portate avanti, ormai da trent'anni, dal CERL, tra i cui meriti vi è appunto l'aver creato un terreno comune per professionisti del libro (bibliotecari, curatori) e ricercatori che hanno nei libri le loro fonti principali.

informatiche²⁷. Allo stesso modo è stato naturale confrontarsi con gli studi condotti sulle edizioni a stampa e più in generale con progetti di censimento annunciati e solo in parte realizzati. Tra le operazioni più interessanti, per l'impostazione generale e per la levatura delle persone coinvolte, è il censimento dei manoscritti delle opere di Bartolo, avviato da Emanuele Casamassima nel 1971 con la pubblicazione del censimento dei manoscritti conservati in Germania; proseguito con la pubblicazione del censimento condotto in Spagna da Antonio García y García e arrestatosi con la preparazione dell'*Iter Italicum*, lasciato da Casamassima allo stadio della raccolta di schede. Quell'ambizioso progetto può ora essere ripreso in una prospettiva diversa, facendo confluire i dati già pubblicati nella base dati, dove possono essere aggiornati, se necessario, e arricchiti di informazioni che il censimento non prevedeva. Al contempo l'opera di analisi e identificazione dei testi, già operata da Casamassima (e prima di lui da Paradisi e Calasso, e insieme a lui da García y García, e poi da quanti hanno continuato a lavorare all'opera di Bartolo, per restare all'esempio citato), costituisce un punto di riferimento per popolare correttamente la base di dati, in particolare per costruire al meglio possibile le schede delle opere, sulla cui lunga tradizione solo gli specialisti possono correttamente informare.

Riuniti nello stesso luogo digitale, i progetti pregressi possono essere ripresi in mano, con le nuove possibilità che la rete consente. ICo non rappresenta certo la prima proposta in tal senso, ma è la prima proposta che ha il focus su testi, che considera insieme manoscritti e a stampa e che ambisce a svilupparsi come un utile strumento di ricerca, prima ancora che luogo di pubblicazione dei dati raccolti. La macchina è progettata per essere in parte personalizzata, per rispondere al meglio alle domande di ricerca poste dall'editor. Se infatti lo scopo è studiare la trasmissione dei testi, la loro registrazione (ad esempio quali testi trascrivere integralmente e quali no) dipenderà dalle scelte dell'editor, che le potrà giustificare utilizzando gli appositi spazi. Quanto spazio riservare alle glosse? Dipende da quali glosse, su quali testi, per quali periodi. Certamente ci sono elementi fondamentali che non possono mancare, ma sulla profondità dell'analisi decide l'editor.

I record sono tutti autoriali, la modalità di citazione è predisposta dalla base dati (con un *cite as*, che si può copiare e incollare) per cui l'autore ha la responsabilità in chiaro delle sue analisi e un modo per vedersi riconosciuto quel lavoro – si sta anche pensando alle modalità di dare una maggiore riconoscibilità scientifica a questo lavoro, con pubblicazioni vere e proprie che lascino al record la sua caratteristica principale, poter essere aggiornato.

Perché è possibile lasciare tanta libertà agli/alle editor? Perché ICo non si sostituisce ai cataloghi, anzi con essi è perfettamente allineato, attraverso un sistema di rinvii che tiene presenti i principali cataloghi online: ISTC e GW per le edizioni incunabile; Edit16, SBN e HPB, sono di preferenza i cataloghi considerati per le edizioni successive, mentre per i manoscritti il riferimento va certamente a *Manuscripta juridica* e a

²⁷ Ci si riferisce naturalmente al censimento dei manoscritti di diritto romano realizzato da Gero Dolezalek e da lui trasformato nella base di dati *Manuscripta juridica* (<<https://manuscripts.rg.mpg.de/>>).

Manus Online – rispetto al quale in realtà si vorrebbe fare di più che un rinvio. Accanto ai cataloghi, sono previsti rinvii diretti agli altri strumenti digitali specializzati, che insistono in tutto o in parte sullo stesso segmento di universo bibliografico: abbiamo già visto i rinvii alle basi di dati sui tipi (TW), sulle marche (Edit16), sulle immagini (15cIllustration e ICONCLASS), sulle filigrane (Bernstein). A queste si devono aggiungere rinvii alle basi di dati sviluppate dal progetto EMOBookTrade: Early Modern Book Privileges e Early Modern Book Prices²⁸; con la prima, in particolare, il legame potrà creare la perfetta corrispondenza con i testi dei privilegi rilevati nelle edizioni, e trascritti, in tutto o in parte, in ICo.

Vanno infine menzionate alcune basi di dati con cui ICo ha stabilito un dialogo più stretto, contribuendo ad alimentarle, direttamente o indirettamente. Indirettamente significa attraverso l'opera degli editor, ciò che vale in particolare per gli esemplari. Se in ICo sono analizzati i testi, dove si dà conto delle specificità degli esemplari e dei possessori? Ovviamente nelle basi di dati dedicate, MEI per gli incunaboli e HPB Provenance per gli esemplari delle edizioni dei secoli successivi. Qui è previsto anche un link diretto verso i record in Manus Iuridica, progetto speciale concepito appunto per riunire nella principale base di dati italiana le descrizioni dei manoscritti giuridici conservati nelle biblioteche italiane o altrove. I legami verso queste basi di dati sono forniti dall'editor in apertura della scheda, in associazione all'esemplare o agli esemplari esaminati. Nel front-end, queste informazioni si leggono in calce alla scheda, in una sezione piccola ma importante, dove sono esplicitati i riferimenti ai cataloghi, la completezza del record e le fonti, i nomi degli editor, il permalink e le istruzioni per la completa citazione del record.

Del dialogo con IUS Illuminatum è stato già detto; resta invece da menzionare il legame diretto con un altro progetto, una prima forma di interoperabilità che ambisce a valorizzare le informazioni prodotte in ICo mettendole a disposizione di un'altra base dati, *Onomasticon. Prosopografia dell'Università degli Studi di Perugia*. Quest'ultima, nata per testimoniare vita e carriere di docenti e studenti, non aveva una sezione per la registrazione delle opere dei maestri, oggi la chiamiamo produzione scientifica. Invece di sviluppare appositamente un ulteriore database bibliografico si è ritenuto di "sfruttare", relativamente ai giuristi, il lavoro già svolto in ICo, dove le loro opere sono via via registrate. La soluzione adottata prevede uno scarico di dati, da ICo ad Onomasticon, della lista delle opere²⁹; da Onomasticon si rinvia ad ICo per recuperare tutte le edizioni delle opere elencate. Il trasferimento dei dati è funzionale, per cui la lista delle opere, che via via in ICo viene incrementata, viene aggiornata ad ogni nuova ricerca effettuata nel database ricevente Onomasticon.

²⁸ I link verso le due basi di dati si recuperano dal sito web del ben noto progetto diretto da Angela Nuovo (<emobooktrade.unimi.it>), di cui un resoconto recente si legge in questa stessa rivista: Andrea Ottone — Erika Squassina, *The Early Modern Book Trade Project: Premises, Objectives, Methodologies and Resources*, «DigItalia. Rivista del digitale nei beni culturali», 17 (2022) n. 2, p. 61-77.

²⁹ Attraverso un'API approntata allo scopo dagli IT responsabili dei due progetti: Luigi Tassarolo, già citato, per ICo e Giulio Quaresima per Onomasticon.

Conclusioni e prospettive

Nel 2024 ICo cambia veste: un nuovo logo, realizzato da una vera professionista sostituisce quello utilizzato per i primi anni del progetto. Prodotto in totale economia, il primo logo di ICo (ancora visibile nelle immagini allegate a questo contributo) è stato progettato per esprimere un'idea semplice che è stata mantenuta: una stilizzazione del segno di paragrafo, che si trova a ritmare, in rosso e blu, le pagine dei manoscritti e delle edizioni più antiche. Nella nuova realizzazione, il disegno è ulteriormente semplificato e riflette bene la speranza che lo strumento si riveli veramente utile a districare la complessa rete di relazioni che esiste tra le varie entità bibliografiche soprattutto in ambito giuridico. Perché questo avvenga è essenziale che semplice si riveli l'uso della base dati da parte degli editor, ed è questo uno degli obiettivi perseguiti nel progettare. La pubblicazione della nuova release – oltre al logo è stata ridisegnata anche la homepage – marca la fine di un primo periodo, durante il quale un piccolo gruppo di giovani studiosi e studiosi, soprattutto dottorande e dottorandi, hanno sperimentato ICo per le loro ricerche. Esigenze specifiche hanno stimolato l'implementazione di nuove funzionalità e ora il disegno della base dati può dirsi completo. I prossimi sforzi andranno quindi in direzione della comunicazione, con una cura maggiore delle pagine web che corredano la base dati, così da agevolarne la conoscenza e l'uso. Si sta studiando una soluzione che consenta al lungo e corposo lavoro di analisi che talune descrizioni richiedono di essere riconosciuto alla stregua di una pubblicazione, ma ogni record è già ora perfettamente citabile. L'obiettivo di lungo periodo è che la base dati possa riuscire nel suo intento principale di proporsi come strumento di ricerca agile ed effettivamente efficace, per far ripartire ricerche interrotte, iniziarne di nuove, porsi come luogo di incontro e di scambio per ricercatrici e ricercatori impegnati a ragionare sulle trasmissioni del testo in un ambito tanto specifico quanto ampio. Non è stato ancora sottolineato, e sarà utile farlo in chiusura, che in ICo descrizioni ed etichette sono in inglese; ciò ha lo scopo di facilitare la cooperazione a livello internazionale, oltre ad essere un omaggio alle origini della base dati.

"Towards a corpus of texts of the ius commune from manuscript to print", so reads the complement of the title of IVS Commune online, the digital system which forms the object of the present article. Such title serves to clarify the purpose of the system: the reconstruction of a corpus of texts notoriously very rich and complex, and very difficult to tackle. The first paragraph is meant to illustrate the terms of such corpus, which is kept wide. Once defined the object of the research and the context in which it was conceived, the article develops by illustrating the digital tool, the underpinning rationale, and main features; special attention is dedicated to aspects of interoperability, data reuse, and, finally, cooperation. Particularly important is considered the cooperation between editors who work in different contexts (academia and holding institutions). A final paragraph serves to illustrate the most recent developments of the tool and to present future perspectives, for the near future and in the long-term.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Una base dati per lo studio del sistema del privilegio librario nell'Europa moderna

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00098

Marius Buning — Alberto José Campillo Pardo — Alica-Nana Citron — Andrea Ottone¹
University of Oslo. Department of Archaeology, Conservation and History

*Il presente saggio discute le premesse teoriche, le funzionalità operative e le prospettive divulgative di una banca dati inerente alla storia del privilegio librario nella prima età moderna. La base dati è stata sviluppata all'interno del progetto *Before Copyright*, finanziato dal Consiglio Europeo delle Ricerche e strutturato presso il Dipartimento di archeologia, conservazione e storia dell'Università di Oslo. Dilatando lo spettro di ricerca dalla storia del libro, di cui il privilegio librario è stato sino ad ora naturale area di pertinenza, il progetto si propone di toccare ambiti quali la storia intellettuale, la storia giuridica e la teoria del diritto. Il presupposto teorico è che il privilegio librario, oltre ad essere stato uno strumento di promozione commerciale, sia stato anche il mezzo attraverso cui gli stati territoriali interferivano nelle dinamiche di filtraggio e sviluppo del sapere. Al fine di individuare un modello interpretativo stabile, fondato sullo spoglio di una vasta messe di fonti storiche, è stato creato un database relazionale. Questo avrà lo scopo di frammentare e riaggregare secondo propositi logici dinamici la molteplicità di dati bibliografici, geografici e biografici derivati da una grande quantità di privilegi emessi nei diversi stati europei per un periodo cronologico compreso fra il 1470 e il 1800 circa, quando cioè il sistema del privilegio venne progressivamente sostituito dal più moderno sistema del diritto d'autore.*

1. Premesse e propositi del progetto *Before Copyright*

La rilevanza del privilegio librario a supporto della nascente industria tipografica rinascimentale non è sfuggita agli storici del libro. Il privilegio e il sistema paneuropeo che si sviluppò sin dalla fine del XV secolo intorno ad esso fornì un

¹ Il presente saggio è stato composto nella seguente modalità: Andrea Ottone ha redatto i paragrafi 1-3, Marius Buning ha redatto in inglese il paragrafo 4, tradotto poi da Andrea Ottone. I presupposti teorici e scientifici del progetto sono stati gettati da Marius Buning, Prime Investigator del progetto *Before Copyright*. Il progetto è finanziato dall'Unione Europea (ERC, BE4COPY, 101042034). Le opinioni espresse in questo articolo sono, tuttavia, formulate esclusivamente dagli autori e non riflettono necessariamente quelle dell'Unione Europea o dell'European Research Council. Né l'Unione Europea né l'istituzione patrocinate debbono essere perciò ritenute responsabili per queste. Il database *Before Copyright* è stato sviluppato da Andrea Ottone in qualità di digital manager del progetto da marzo 2023 a febbraio 2024. L'assetto strutturale del database ha visto anche il contributo con-

supporto giuridico a protezione delle industrie locali calmierando la competizione e promuovendo l'innovazione².

Principali attori di questo sistema furono le istituzioni locali o nazionali da una parte e gli editori dall'altra. Per questa ragione, gli studi sul sistema del privilegio librario si sono mossi dapprima nell'ambito della storia del libro; solo occasionalmente hanno riguardato la storia e la teoria del diritto e, ancor più raramente, altri ambiti disciplinari quali la storia intellettuale e l'epistemologia. La storia della censura libraria, cui il privilegio di stampa è spesso strettamente legato, non ha condiviso questa stessa sorte, avendo ricevuto invece attenzioni da parte di settori trasversali che includono, oltre alla storia del cristianesimo e delle istituzioni, anche la storia culturale e intellettuale, come dimostra l'interesse degli storici della filosofia e della scienza nei confronti delle avverse sorti di molti esponenti del nuovo pensiero. Tuttavia, proprio come la censura libraria, anche il privilegio librario si configurava come mezzo di filtraggio istituzionale del sapere. Sebbene ad essere protetto, almeno in origine, fosse il valore intrinseco del libro come merce, è tuttavia indubbio che questa merce acquisiva valore di mercato anzitutto in ragione dell'appetibilità del contenuto che questa veicolava. Di riflesso, quindi, le istituzioni preposte all'erogazione dei privilegi librari divenivano parte attiva nella promozione di determinati contenuti e forme del sapere.

La qualità di un testo, la rilevanza dei contenuti ma anche la notorietà di un autore o il valore strategico della disciplina rappresentata da un'opera divenivano tutti fattori influenti nella valutazione dell'opportunità di proteggere un'edizione con un privilegio. Il privilegio librario rivela dunque le aspettative commerciali che un'intera comunità, attraverso le proprie istituzioni, riponeva in un determinato testo. Le petizioni di privilegio, provenienti da autori o da editori, ponevano al centro delle motivazioni il timore di iniziative rapaci di ristampa; la concessione del privilegio riconosceva implicitamente il rischio di una competizione sleale sul prodotto in oggetto. L'ottenimento di un privilegio, spesso orgogliosamente ostentato sul frontespizio, era non solo un mezzo di protezione legale ma anche uno strumento di promozione commerciale. Il privilegio, infatti, evidenziava il valore aggiunto di un'opera in ragione del possibile interesse che questa avrebbe potuto suscitare presso alcuni competitori³.

tinuo di Alberto Campillo Pardo, che ha assunto il ruolo di nuovo digital manager da marzo 2024, e Alica-Nana Citron: questi ultimi sono inclusi tra gli autori in quanto i loro interventi e consigli hanno contribuito alla strutturazione del database qui presentato. Alberto Campillo Pardo ha inoltre supervisionato il dataset e l'elaborazione dei grafici in Figura 7 e 8. Si ringraziano anche Pim van Bree e Geert Kessels di LAB1100 per l'aiuto fornito nel perfezionare il modello del database in ambiente Nodegoat. Si coglie anche l'occasione per ringraziare Flavia Bruni per il suo prezioso aiuto in fase di revisione testuale.

² Angela Nuovo, *Le politiche legislative sulla stampa in età moderna*, in: *Privilegi librari nell'Italia del Rinascimento*, a cura di E. Squassina e A. Ottone, Milano: FrancoAngeli, 2019, p. 9-15; Id., *The Book Trade in the Italian Renaissance*, Leiden: Brill, 2013, p. 195-212.

³ Marius Buning, *Privileging the common good: the moral economy of printing privileges in the seventeenth-century Dutch Republic*, in: *Buying and selling. The business of books in early modern Europe*, a cura di S. Graheli, Leiden: Brill, 2019, p. 88-108.

Il privilegio di stampa era inoltre una forma di sponsorizzazione ufficiale da parte di autorità territoriali o religiose. Le autorità emittenti si rivolgevano al richiedente con atteggiamento paternalistico, apprezzandone lo sforzo, il valore e la lealtà personale, per riconoscere poi il significato che il contributo intellettuale proposto avrebbe avuto per l'intera comunità e concedendo dunque il privilegio come atto di benevolenza particolare: nei privilegi papali il beneficiario veniva definito "figlio prediletto"⁴. Similmente, nelle corti laiche la concessione di un privilegio veniva enunciata in una forma non scevra di paternalismo, come in questo privilegio concesso dal Senato Veneto all'editore Michele Tramezino nel 1564:

«Che al fidel nostro Michiel Tramezino sia concesso che niuno altro che lui o chi havrà causa da lui non possa stampar, né far stampar in questa città, né in alcuna altra città o luogo della Signoria nostra, né altrove stampata vender, l'opera intitolata Li essercitii, over meditazioni de i sette giorni della settimana di frate Luigi di Granata, et I sette essercitii di santa Geltruda, et La quinta, et sexta parte del cavaliere Sferamundi, per lo spatio di anni vinti prossimi, sotto pena a i contraffattori di perder tutte le opere da loro stampate, et ducati dieci per cadauna di quelle, uno terzo della quale habbia ad esser del magistrato che farà l'essecutione, uno terzo dell'arsenal, et l'altro di lui supplicante, essendo obligato di osservare quanto è disposto in materia di stampe»⁵.

Difficile non vedere, in questo dialogo fra attori del mondo del libro e istituzioni territoriali, un'intersezione fra commercio, cultura e politica, ove la politica interferiva con la cultura in ragione del commercio, promuovendo, verrebbe da dire privilegiando, determinate fatiche dell'intelletto a discapito di altre.

Il sistema del privilegio era, in ultima analisi, una forma di negoziazione fra privati cittadini e istituzioni, come rivelato dall'etimologia del termine: *privata lex*. Il privilegio librario ricopriva, forse in ragione della prominenza e della strategicità dell'industria libraria occidentale, un ruolo estremamente attivo all'interno di questo sistema. Questo genere di negoziazione privata, che spesso coinvolgeva diversi intermediari in forme clientelari tipiche dell'antico regime, sottopone i parametri di definizione dei singoli privilegi ad un interessante dinamismo. Tracciare le forme di questo dinamismo e lo sviluppo di modelli riconoscibili nello spazio e nel tempo, desumendo da questi eventuali epifenomeni corrispondenti è il proposito del progetto *Before Copyright*.

Con l'intento di coprire un ampio spettro geografico, ovvero lo spazio europeo in cui la civiltà tipografica era attiva, unitamente al proposito di misurare l'evoluzione

⁴ Si veda ad esempio Archivio Apostolico Vaticano, Armarium XLII, volume 29, f. 239r, cit. in Jane Ginsburg, *Proto-proprietà letteraria ed artistica: i privilegi di stampa papali nel XVI secolo*, in: *Privilegi librari nell'Italia del Rinascimento*, cit., p. 122.

⁵ Archivio di Stato di Venezia, Senato, Terra, b. 45, f. 79r-v, cit. in Erica Squassina, *Privilegi librari ed edizioni privilegiate nella Repubblica di Venezia (1527-1565)*, Milano: Milano University Press, 2022, p. 360.

del sistema dei privilegi nell'arco di tre secoli, si è ritenuto necessario l'utilizzo degli strumenti digitali, in quanto in grado di accogliere e gestire una copiosa quantità di dati.

2. La base analogica: dati desumibili da un privilegio librario e loro valore scientifico

Di norma, il privilegio librario situa un evento (la protezione di un'opera letteraria, artistica, tecnica o un'innovazione produttiva legata alla stampa) in un definito contesto geografico (i confini di validità, corrispondenti ai limiti giurisdizionali dall'autorità emittente) e in un determinato periodo di tempo (approssimativamente la durata del privilegio), chiamando in causa determinati attori, generalmente il richiedente e il beneficiario (quando questi non erano la stessa persona, o gruppi di persone)⁶. Tuttavia, ampliando il terreno di osservazione, si può notare come fossero coinvolti molteplici altri elementi. Dall'esempio del 1564 già riportato si evince una ricca messe di dati: un privilegio poteva riguardare molteplici opere letterarie, come nel caso citato, in cui erano tre. Ogni opera poteva poi chiamare in causa più autori, occasionalmente connessi ad altri attori come commentatori, curatori e traduttori. Nell'opera privilegiata potevano inoltre essere presenti illustrazioni, a loro volta collegate ad altre responsabilità. Lo spazio di azione di un privilegio corrispondeva generalmente a quello rappresentato dall'area giurisdizionale dell'autorità emittente: un cittadino o residente di Venezia richiedeva un privilegio valido all'interno del Dominio della Repubblica; tuttavia, i rapporti spaziali potevano essere più articolati nel caso di istituzioni con più ampia giurisdizione⁷. La corte imperiale, ad esempio, poteva concedere privilegi validi all'interno del frastagliato territorio del Sacro Romano Impero anche a cittadini stranieri. Nel 1569 l'imperatore Massimiliano II concesse un privilegio quinquennale al libraio veneziano Marco Antonio de Amadori per gli *Instituta canonica cum glossis* del giurista perugino Giovanni Paolo Lancellotti e le *Conclusiones et opiniones compositae* di un altro giurista, stavolta romano, Antonio Gabrieli⁸: in ragione di tutti gli attori in gioco, le intersezioni geografiche si complicarono notevolmente.

⁶ La definizione dei dati temporali legati a un privilegio è in realtà soggetta a variabili. In alcuni Stati si prevedeva che la durata del privilegio fosse definita dall'ultimazione della stampa di un'opera, rendendo quindi la data di concessione del privilegio solamente indicativa. Si dava inoltre il caso che in alcuni sistemi giuridici il privilegio scadesse qualora l'ultimazione della stampa dell'opera protetta non venisse ultimata entro un anno dalla concessione: cfr. A. Nuovo, *The book trade*, cit., p. 212.

⁷ A Venezia raramente si concedevano privilegi per iniziative editoriali condotte al di fuori dei confini della Serenissima, limitando quindi la relazione spaziale di un privilegio (cfr. A. Nuovo, *The book trade*, cit., p. 209-210); era quindi usuale che un cittadino o residente veneziano richiedesse un privilegio valido nei territori della Repubblica. Per una nota eccezione si veda Erika Squassina, *La protezione del Furioso: Ariosto e il sistema dei privilegi in Italia*, «Bibliotheca.it», 6 (2017), n. 1, p. 9-38, DOI <<https://doi.org/10.6092/issn.2283-9364/7024>>.

⁸ Österreichische Staatsarchiv, AT-OeStA/HHStA RHR Impressoria 1/2-20. Si ringrazia Alica-Nana Citron per la segnalazione di questo esempio.

Tornando all'esempio veneziano proposto prima, si noterà che un'ulteriore articolazione riguarda l'apparato delle pene previste in caso di infrazione. Nel caso specifico, la pena pecuniaria era da destinarsi a tre enti, beneficiari di un terzo ciascuno della sanzione. L'ammontare di quest'ultima, come anche la sua distribuzione fra i destinatari e l'identità degli stessi, poteva variare per periodi temporali e realtà statuali differenti, offrendo un potenziale di interessanti casistiche e modelli interpretativi derivanti. In definitiva, la gestione delle sanzioni legate all'eventuale infrazione di un privilegio intersecava l'interesse più strettamente politico della realtà sociale in cui il privilegio veniva emesso. Nel caso del privilegio concesso a Michiel Tramezino, uno dei dividendi della sanzione era destinato all'Arsenale di Venezia, istituzione strategica per la condotta commerciale e militare della Repubblica. Dietro il variare dei destinatari d'una sanzione come dietro il variare del loro peso si leggono in controluce epifenomeni che legano insieme priorità culturali, sociali e istituzionali⁹.

Un'altra variabile di estrema rilevanza riguarda il valore culturale espresso dalle opere privilegiate. Il genere letterario lega a stretto giro le vicissitudini dei privilegi librari alla storia intellettuale della prima età moderna. Nell'indicare il potenziale pubblico di lettori cui l'opera era destinata, il genere letterario connette il privilegio alla realtà viva della comunità intellettuale in cui questo era immerso. Nel prediligere un'opera o un genere, le istituzioni supportavano determinati settori del sapere a discapito di altri. Questa modalità di supporto e promozione era ulteriormente manifestata dalla gestione delle sanzioni: quelle più aspre esprimevano infatti una specifica valutazione sul valore commerciale ma anche strategico e intellettuale dell'opera che si intendeva proteggere. Differenti modelli di protezione potrebbero quindi evidenziarsi in base ai campi del sapere rappresentati, seppure approssimativamente, dai diversi generi letterari delle opere privilegiate.

3. La base dati digitale: frammentazione e riaggregazione dei dati desunti

Al fine di gestire la molteplicità di dati desumibili dalle fonti inerenti al sistema del privilegio librario europeo, la scelta del progetto *Before Copyright* è caduta sul software Nodegoat, programma *freeware* sviluppato dalla società olandese LAB1100, in ragione della versatilità della piattaforma nella strutturazione di database relazionali, della possibilità di gestire dati biografici e geografici in prospettiva storica e, in ultimo, delle molteplici funzionalità di creazione e visualizzazione di reti sociali¹⁰.

⁹ Sulla politica di assegnazione delle sanzioni si veda l'analisi offerta per il caso veneziano in Erika Squassina, *The protection of the printing industry in Venice (1560-1580)*, in: *Competition in the European Book Market: Prices and privileges (fifteenth-seventeenth centuries)*, a cura di A. Nuovo, J. Proot, D. Booton, Antwerp: Vereniging van Antwerpse Bibliofielen, 2023, p. 235-251.

¹⁰ <https://nodegoat.net/>.

Nel presentare il back-end della base dati BE4COPY è opportuno partire dalla maschera principale di compilazione dati, ovvero la tabella *Privilege* (Fig. 1). Come mostra lo schema logico del back-end, la tabella *Privilege* dà accesso a tutte le tabelle derivate (Fig. 2).

Figura 1. Back-end del database BE4COPY: tabella principale Privilege

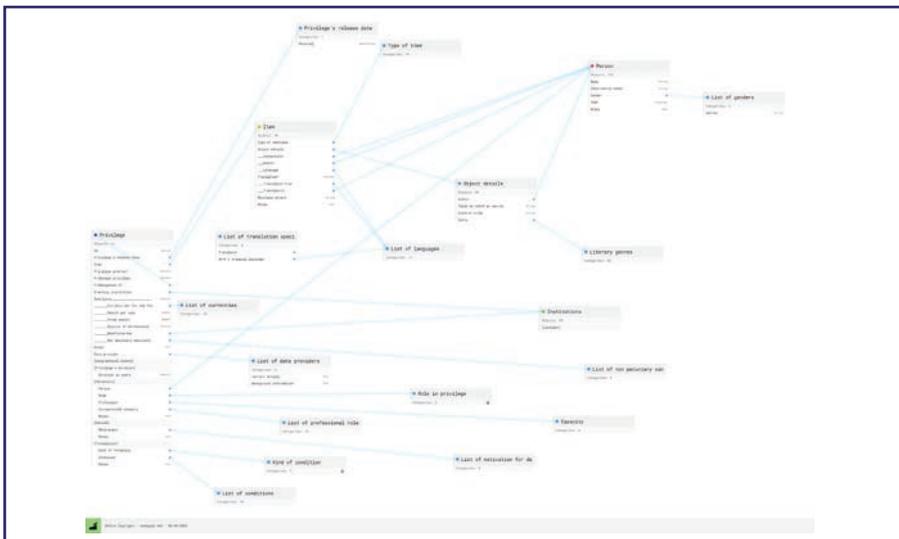


Figura 2. Modello generale della base dati BE4COPY

Fra queste, la tabella *Item*, insieme all'altra tabella derivata *Object details* e alla scheda *Person*, danno modo di inserire una molteplicità di dettagli bibliografici (cfr. Fig. 3-4).

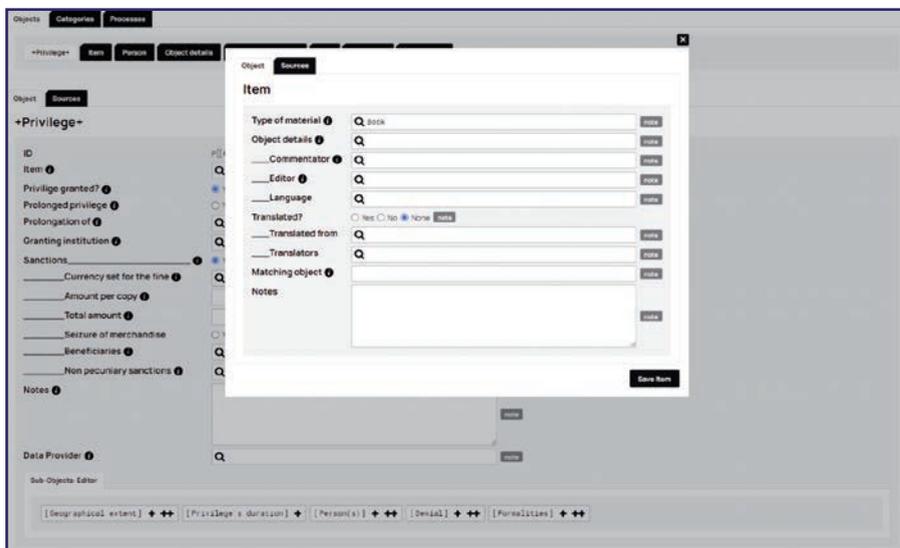


Figura 3. Back-end del database BE4COPY: tabella *Item*

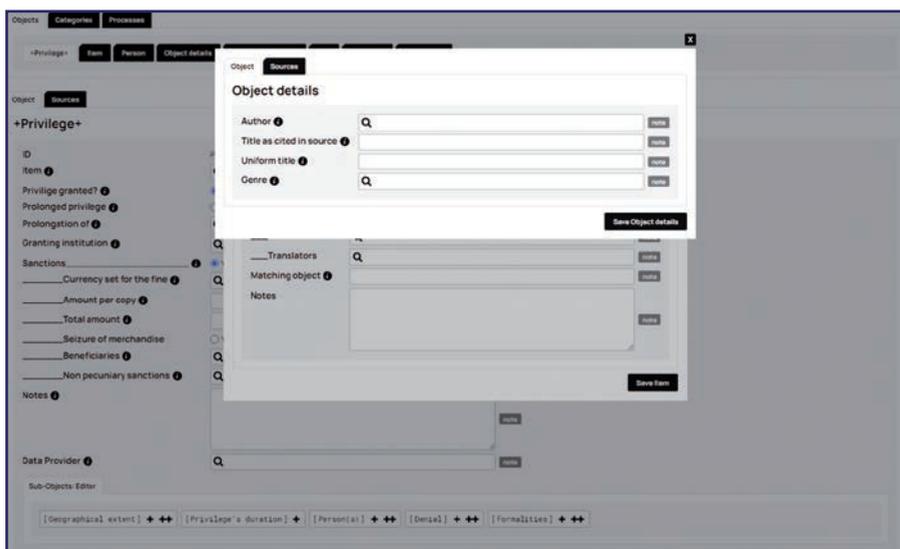


Figura 4. Back-end del database BE4COPY: tabella *Object details*

L'inserimento dei dati avviene percorrendo a ritroso i tre livelli enunciati. Attraverso la tabella *Person*, compilando il campo *VIAF* i dettagli inerenti all'autore sono attinti direttamente dal database Virtual International Authority File¹¹.

¹¹ <https://viaf.org/>.

All'interno della scheda si possono inserire, oltre al nome in forma accettata, forme di rinvio, genere (*Male*, *Female* e *Unknown* nel caso, ad esempio, di entità collettive) ed eventuali note biografiche aggiuntive (Fig. 5-6). Questi dati verranno

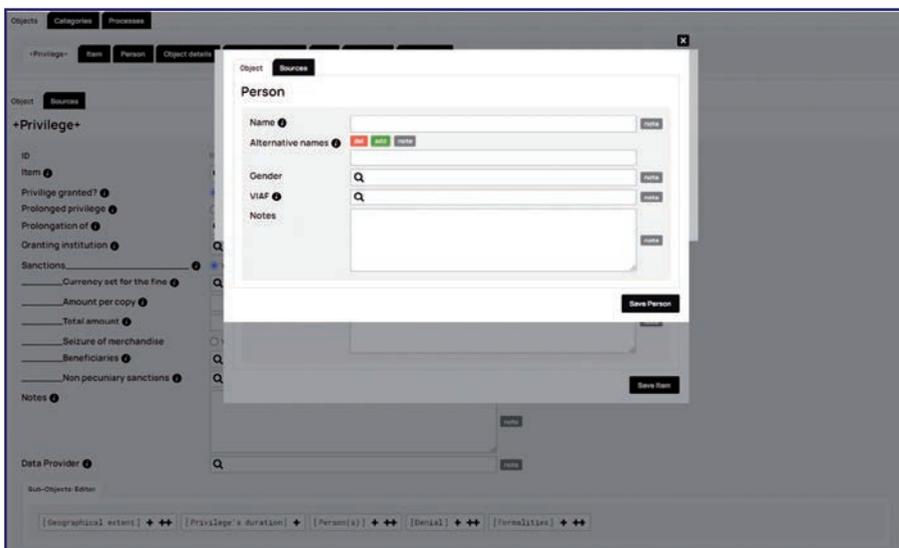


Figura 5. Back-end del database BE4COPY: tabella Person

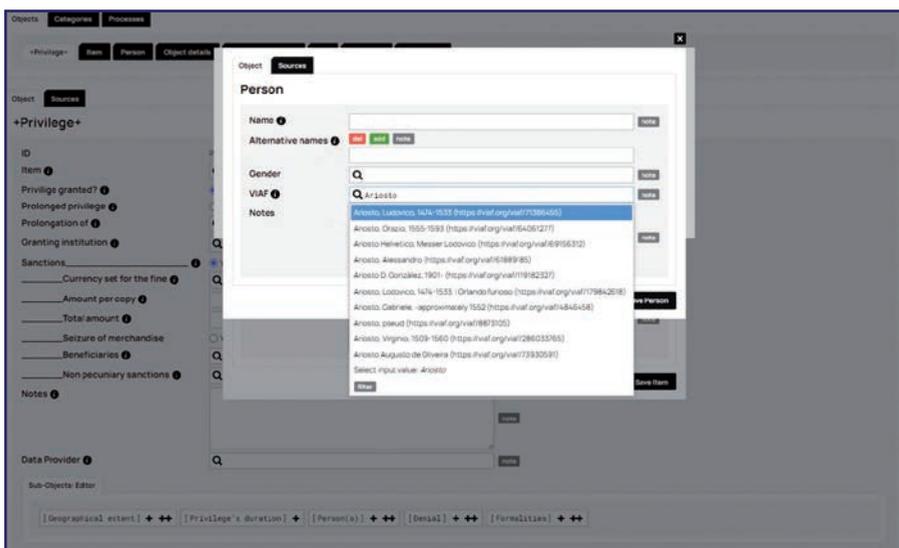


Figura 6. Back-end del database BE4COPY: modalità di ingestione dal database Virtual International Authority File

integrati nel campo *Author*, che è predisposto per contenere molteplici record in caso di opere con autore multiplo. Compilato il campo autore, l'inserimento prosegue tramite la maschera *Object details*, in cui si può fornire il titolo come citato

nella fonte d'archivio o il titolo uniforme; vista la rilevanza per gli scopi analitici del progetto, il genere dell'opera deve essere obbligatoriamente inserito, selezionando da una tendina una sola fra molteplici opzioni basate sulla tassonomia elaborata da Jacques-Charles Brunet¹². Nella maschera *Item*, dopo aver completato la scheda di dettaglio definendo il tipo di oggetto privilegiato (*Book, Broadsheet, Cartography, Globe* ecc.), è possibile fornire dettagli sul commentatore o curatore (*Editor*) attingendo ancora una volta alla tabella *Person* già illustrata (Fig. 4), la lingua o le lingue (non applicabile nel caso di oggetti grafici, come illustrazioni e mappamondi, o partiture musicali prive di testo); nel caso di traduzioni, è possibile specificare la lingua originale e il traduttore (o i traduttori) compilando la corrispondente scheda *Person*. Un campo opzionale (*Matching object*) è riservato all'identificazione dell'opera tramite repertori esterni, in particolare inserendo permalink di record bibliografici.

Nella maschera principale *Privilege* è poi possibile specificare l'esito della petizione (ove la negazione, soprattutto se motivata, è di notevole interesse nel quadro analitico del progetto), o se il privilegio fosse il prolungamento di uno precedentemente elargito. Proseguendo si può fornire notizia riguardo l'autorità emittente, i dettagli sulle sanzioni pecuniarie comminate e sugli eventuali beneficiari.

Nel quadro funzionale della piattaforma Nodegoat gli oggetti subordinati (*Sub objects*) inerenti all'oggetto principale (*Privilege*) permettono di gestire le attribuzioni cronologiche e geografiche dei dati di interesse connessi all'oggetto stesso, nello specifico (Fig. 1): *Geographic extent* (finalizzato a definire l'area di influenza del privilegio librario in oggetto), *Privilege duration* (tesa a definire la durata del privilegio) e *Person(s)* (destinato a gestire la determinazione biografica del richiedente o beneficiario di un privilegio). Sempre nell'area dei *sub objects* sono presenti le tabelle *Denial* e *Formalities*: la prima è destinata a fornire dettagli sull'eventuale diniego di un privilegio e la seconda a fornire elementi inerenti agli eventuali obblighi stipulati dal privilegio per garantirne la validità.

Per la gestione dei dati geografici, Nodegoat mette a disposizione un database di 137.514 luoghi pre-settati, offrendo ai singoli progetti la possibilità di aggiungerne altri nel proprio spazio gestionale all'interno della piattaforma. Quest'ultima, inoltre, integra lo standard GeoJSON permettendo una buona duttilità nel produrre schemi di visualizzazione dei dati. La funzione viene utilizzata per la schematizzazione visuale delle aree di influenza dei privilegi librari. La visualizzazione cronogeografica di dati relativi all'età moderna deve necessariamente tenere conto dei possibili cambiamenti di denominazione o alterazione dei confini territoriali di alcune località. La possibilità di creare geometrie fittizie per rappresentare confini variabili permette di visualizzare le aree di estensione dei privilegi e i rapporti di prossi-

¹² Si rimanda indicativamente all'edizione inglese Jacques-Charles Brunet – Donald Bruce McKeon, *The classification system of Jacques-Charles Brunet*, Baton Rouge: Louisiana State University, Graduate School of Library Science, 1976.

mità spaziale o sequenzialità temporale nell'ambito del mercato librario europeo che, in età coloniale, poteva estendersi alle aree transoceaniche. La visualizzazione dei dati inerenti al privilegio librario europeo su una mappa e l'intersezione con elementi cronologici offrono il vantaggio di dare la misura del carattere transnazionale del fenomeno, in particolare per realtà come il Sacro Romano Impero o la Santa Sede, la cui giurisdizione valicava idealmente i confini statuali del reggente (Fig. 7).

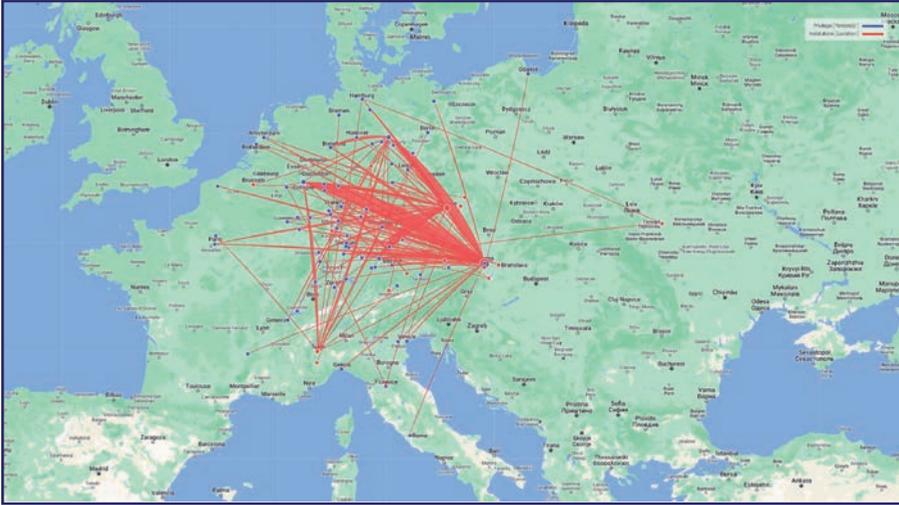


Figura 7. Visualizzazione dei dati geografici inerenti ai privilegi rilasciati dalla corte imperiale nel XVIII secolo

Un'altra funzione interessante di Nodegoat riguarda la modalità di visualizzazione dei collegamenti all'interno di una rete sociale. Ciascun elemento isolato nel processo di frammentazione delle informazioni inerenti a un privilegio (beneficiario, autore, editore, titolo, genere letterario, autorità emittente ecc.) diventa un nodo all'interno di un reticolo di dati interconnessi. I nodi ricorrenti con maggiore frequenza o con un numero più alto di relazioni con altri nodi vengono evidenziati nella mappa relazionale secondo gradienti dimensionali variabili stabiliti manualmente in fase di analisi dei dati (Fig. 8); al contrario, i dati con un minor numero di relazioni sono collocati al margine della rete relazionale come punti periferici. La visualizzazione dell'intero tessuto reticolare generato dal dataset può essere gestita diacronicamente, permettendo di isolare determinati periodi storici e di mostrare la sequenzialità della distribuzione dei dati con il progredire del tempo.

La doppia modalità di visualizzazione dei dati, quella crono-geografica e quella di formazione di reti relazionali nel tempo, permette di studiare l'evoluzione del sistema del privilegio su scala europea da molteplici angolazioni ponendo in luce l'intersezione tra fenomeni sociali, politici e culturali e la gestione dinamica della politica di controllo sulla produzione del sapere.

di accreditamento della proprietà intellettuale sui dati forniti e sul lavoro svolto che si dipana su più livelli. I collaboratori verranno riconosciuti come autori dei dati all'interno del database, mentre al gruppo *Before Copyright* spetterà il compito di uniformare le informazioni al fine di popolare la banca dati in accordo con i propositi scientifici e gli standard editoriali del progetto. Il progetto offre inoltre l'opportunità di pubblicare i dati, unitamente ai risultati analitici raggiunti, come saggi all'interno di miscelanee o come monografie autonome in una collana digitale con revisione fra pari e ad accesso aperto. In entrambi i casi, l'apparato critico può variare in base al livello di approfondimento desiderato, da una semplice presentazione del contesto politico-istituzionale e commerciale della realtà statale che ha prodotto i privilegi presentati a un'analisi più dettagliata dei dati. La pubblicazione è interamente patrocinata dal progetto, che avrà anche cura dell'editing dei dati per la pubblicazione dell'e-book.

Il progetto avrà corso d'opera sotto gli auspici dell'European Research Council fino ad agosto 2027. I ricercatori che vorranno aderire al piano editoriale proposto potranno contattare il gruppo di ricerca utilizzando le informazioni di contatto disponibili presso la pagina web di *Before Copyright*¹³.

This paper discusses the theoretical premises, the operative functions, and the publishing perspectives of a database on early modern printing privileges. This database has been developed within the framework of the "Before Copyright" project, funded by the European Research Council and pursued at the Department of Archaeology, Conservation and History of the University of Oslo. Although initially focused on book history, this project aims to contribute to fields of study such as intellectual history, legal history and legal theory. The theoretical premise is that printing privileges, aside from being an instrument of commercial promotion, also served as means through which territorial states interfered in cultural and intellectual dynamics by filtering whole fields of knowledge. In order to build an interpretative model around this hypothesis, a relational database was created to manage a vast amount of data. The said database will be used to fragment and reaggregate data according to multiple logical and investigative purposes. The dataset, which includes bibliographical, geographical and biographical elements, is derived from a vast number of printing privileges spanning from 1470s to circa 1800. The ante quem terminus is set at a period of time when the printing privilege system began to decline, eventually being replaced by modern copyright law.

¹³ <https://www.hf.uio.no/iakh/english/research/projects/before-copyright/>.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Tecnologie GIS nell'ambito dei beni culturali ecclesiastici¹

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00099

Laura Gavazzi

Ufficio Nazionale per i beni culturali ecclesiastici e l'edilizia di culto (Conferenza Episcopale Italiana)

Claudio Camuto — Vittoria Rebuttini — Stefania Ricci

Centro Servizi per i beni culturali ecclesiastici (Conferenza Episcopale Italiana)

L'Ufficio Nazionale per i beni culturali ecclesiastici e l'edilizia di culto della CEI ha avviato un progetto per la cartografia dei confini delle diocesi italiane, fondamentale per vari progetti come il Censimento Chiese e la Manutenzione programmata. I Sistemi Informativi Geografici offrono un modo intuitivo e immediato di visualizzare informazioni, permettendo di intersecare dati provenienti da fonti diverse e di condurre analisi dettagliate. Questi sistemi consentono inoltre di aggiungere diversi livelli informativi su una mappa, facilitando la comprensione e l'analisi dei dati.

Si rende inoltre conto dell'utilizzo di informazioni del Catasto Gregoriano e quelle provenienti dai Censimenti delle Chiese tramite tecnologie GIS, fornendo nuovi servizi alle diocesi per la gestione del patrimonio ecclesiastico e in particolare per analizzare il patrimonio ecclesiastico della Diocesi di Bologna.

1. Cartografia Confini Diocesi Italiane

Molti progetti promossi dall'Ufficio Nazionale per i beni culturali ecclesiastici e l'edilizia di culto (BCE) della Conferenza Episcopale Italiana (CEI), quali Censimento Chiese², Censimento Immobili, Manutenzione programmata e tutti quelli che necessitano di rappresentazione di progetti su mappa, hanno l'esigenza e l'utilità di disporre di una riproduzione "esatta" su cartografia dei confini delle diocesi italiane.

L'Ufficio BCE insieme al Servizio Informatico CEI (SICEI)/IDS&Unitel ha favorito un lavoro di verifica di fattibilità sugli strumenti tecnologici e il metodo possibile, coinvolgendo il gruppo di lavoro che si coagula attorno alla rivista «In_Bo. Ricerche e progetti per il territorio, la città e l'architettura» mediante un rapporto convenzionato con il Dipartimento di Architettura e i Dipartimenti di Ingegneria

¹ Il contributo ha le seguenti referenze: 1° capitolo Laura Gavazzi; 2° capitolo Claudio Camuto; 3° capitolo Stefania Ricci; 4° capitolo Vittoria Rebuttini.

² Il Censimento Chiese si riferisce al Progetto di schedatura degli edifici di culto mentre Censimento Immobili è un software di gestione immobiliare e manutenzione programmata di edifici di proprietà ecclesiastica, <<https://bce.chiesacattolica.it/progetto-censimento-chiese/>>.

Civile, Chimica, Ambientale e dei Materiali (DICAM) dell'Alma Mater Studiorum Università di Bologna.

Prima di estendere il lavoro a tutte le diocesi d'Italia, è stato sperimentato solo su due diocesi (Bologna e Ravenna) per capire cosa avrebbe comportato il coinvolgimento generale.

Già dai primi riscontri è emerso un lavoro utile agli Uffici tecnici e alle diocesi stesse che sono soggette a continue soppressioni e accorpamenti di Enti. Il sistema a regime prevede inoltre anche la gestione cartografica di eventi, come ad esempio la rappresentazione in caso di emergenze quali rischio sismico, incendio, idrogeologico ecc.

Lo studio è stato eseguito disegnando confini vettorializzati³.

Si tratta di un processo che restituisce un dato con una precisione accettabile, il metodo si basa su due presupposti indispensabili:

- l'esatta posizione delle chiese (dato che deriva per la maggior parte dal Censimento delle Chiese delle Diocesi Italiane);
- una carta dei confini diocesani, recente se possibile.

Tracciati i confini diocesani, durante la fase di realizzazione di tutto il progetto, sono stati verificati dalle diocesi italiane.

Da questa base consegnata all'Ufficio BCE della CEI, che possiamo definire di "livello 0", ci troviamo già nella fase successiva, quella di aggiornamento del processo con l'inserimento di decreti recenti che sono intervenuti a modificare alcuni confini diocesani: siamo già nella fase di aggiornamento o modifica.

2. GIS un nuovo modo di fruire le informazioni

Le tecnologie Geographical Information System (GIS) si pongono come obiettivo quello di mostrare in maniera immediata e intuitiva le informazioni che servono all'utilizzatore in un determinato contesto attraverso un approccio grafico che risulta facilmente comprensibile e già focalizzato sulle informazioni che si vogliono rappresentare per renderle più agevoli rispetto ad altre visualizzazioni "tradizionali" sotto forma di elenco. Le informazioni poste su mappa presentano un dato, quello della localizzazione, che permette di incrociare informazioni provenienti da fonti profondamente diverse fra loro sia per struttura che per contenuti senza la necessità di dover realizzare complessi meccanismi di associazione. È quindi estremamente facile intersecare informazioni utili e successivamente effettuare delle analisi su di esse o far risaltare determinati aspetti di un contesto. Per chiarire meglio faremo qualche esempio: supponiamo di avere una serie di immobili di proprietà in varie zone del territorio e di averli rappresentati sottoforma di *puntuali* su mappa, a questa mappa che già di per sé fornisce un'informazione utile a chi la consulta aggiungiamo uno strato (*layer* nella terminologia GIS) con l'informazione

³ Luigi Bartolomei – Federica Fuligni, *BeWeB e Georeferenziazione dei confini delle diocesi italiane*, in: *BeWeB 2020. Vent'anni del portale*, Roma: Gangemi, 2020, p. 171-174.

sul *grado di disagio sociale* del territorio, ovvero un indicatore numerico del punteggio assegnato ad ogni zona d'Italia basandosi su indicatori dell'Istituto nazionale di statistica (ISTAT)⁴ quali la percentuale di crimini, la dispersione scolastica ecc.; se "coloriamo" queste zone in base al punteggio partendo da un verde acceso fino ad arrivare a rosso per le zone a maggior rischio otteniamo sin da subito un'informazione: ovvero riusciamo istantaneamente ad avere un'idea di quali immobili si trovano in zone più o meno critiche e quindi, nell'ipotesi di doverne gestire la compravendita, siano più facilmente affittabili o vendibili rispetto ad altri. Se avessimo voluto ottenere lo stesso risultato con degli elenchi tradizionali avremmo fatto molta più "fatica" per correlare fra loro le due informazioni; sarebbe stato necessario, ad esempio, ricavare di ogni immobile dall'indirizzo la zona censuaria in cui si trova e solo allora da elenco ricavare i valori corrispondenti di disagio e poi ordinarli per *grado di disagio*. Un altro esempio di ambito di applicazione è quello della manutenzione e rischio degli immobili, nuovamente mettendo sulla stessa mappa i propri immobili e gli indicatori di rischio sismico, idrico e frane messi liberamente a disposizione da INGV⁵ e ISPRA⁶ è possibile individuare gli immobili che necessitano con maggiore urgenza di interventi di manutenzione o messa in sicurezza; in questo caso la mappa ci aiuta anche ad identificare immobili vicini ma in zone a minor rischio in cui potrebbero essere spostati i beni. Una delle caratteristiche chiave dei sistemi GIS è quindi la possibilità di aggiungere diversi livelli informativi sulla stessa mappa, ogni livello presenta delle proprie informazioni che possono essere combinate con quelle degli altri livelli per ottenere rappresentazioni utili per un determinato obiettivo, potremmo dire per rispondere ad una o più domande. Rispetto ad altre "mappe" a cui siamo abituati al giorno d'oggi (come quelle di Google Maps o dei vari navigatori d'auto) in questi sistemi l'utente può in maniera semplice intervenire su cosa visualizzare e come visualizzarlo oltre ad aggiungere altre informazioni rispetto a quelle previste da chi realizza la mappa. Tali sistemi permettono, infatti, di assegnare stili diversi agli elementi in base alle informazioni; si potrebbe quindi, ad esempio, sull'indicatore del rischio sismico colorare le varie zone del territorio italiano di rosso, arancio, giallo e verde a seconda del grado di rischio così senza neanche aprire il dettaglio si vedrebbe subito sia la distribuzione del rischio nel territorio che il grado di rischio dei singoli immobili. Gli immobili a loro volta potrebbero essere rappresentati con pallini più o meno grandi a seconda del numero di residenti, essere colorati di rosso o di verde a seconda se sia stata controllata o meno la loro stabilità ecc. I sistemi GIS, inoltre, permettono di effettuare analisi e elaborazioni sui dati al fine di generare nuove informazioni; ad esempio data una "sagoma" che rappresenta un territorio (come un confine comunale) il sistema è in grado di calcolarne la superficie in chilometri

⁴ Istituto nazionale di statistica, <<https://www.istat.it/it/archivio/222527>>.

⁵ Istituto nazionale di geofisica e vulcanologia, <<https://www.ingv.it/>>.

⁶ Istituto superiore per la protezione e la ricerca ambientale, <<https://www.isprambiente.gov.it/it>>.

quadrati oppure è possibile unire un *layer* dotato di informazioni cartografiche con altre informazioni non geografiche, ad esempio si potrebbero unire le sagome dei comuni italiani con le altre informazioni statistiche fornite da ISTAT (residenti, aziende, famiglie ecc.). Oltre a quanto detto, molti sistemi GIS come, ad esempio, ESRI ArcGIS⁷ sono dotati di una serie di tool per realizzare delle mini-applicazioni *GIS-based* funzionanti e adatte sia a dispositivi mobili che pc, le quali hanno menu, grafici, filtri, elenchi che interagiscono con le mappe create. Si potrebbe ad esempio andare a indicare il numero di immobili presenti su mappa e con un grafico a torta suddividerli in base al loro stato di conservazione, filtrando sul grafico a torta e/o sulla mappa si possono sovrapporre le due informazioni e quindi disegnando un'area libera visualizzare quanti immobili vi sono in quell'area e il loro stato, aggiungendo il filtro sul grafico si può sapere quanti immobili in quell'area sono in stato rudere. Aggiungendo all'applicazione anche un elenco si può leggere quali sono gli indirizzi e i proprietari degli immobili filtrati. Per realizzare queste applicazioni non è richiesta nessuna competenza di sviluppo/programmazione, il software permette di indicare attraverso lo strumento di creazione cruscotti e applicazioni i dati da visualizzare, come mostrarli e le azioni da intraprendere (ad esempio: al click su un elemento del grafico a torta filtra la mappa sulle voci corrispondenti). Nel seguito di quest'articolo verrà mostrato un esempio concreto di applicazione GIS realizzata per uno specifico ambito.

3. Catasto Gregoriano, Censimento Chiese e GIS

Nel Portale BeWeB⁸ l'ufficio BCE mostra il lavoro di censimento sistematico del patrimonio storico e artistico, architettonico, archivistico e librario portato avanti dalle diocesi italiane e dagli istituti culturali ecclesiastici sui beni di loro proprietà. Con particolare riferimento ai Beni architettonici ad oggi ne risultano 66.615 pubblicati su BeWeB.

La sinergia di BCE con il SICEI ha consentito un costante aggiornamento delle tecnologie IT in uso; l'evoluzione degli ultimi anni ha portato ad erogare i servizi in Cloud (applicazioni software e banche dati), adozione di tecnologie GIS (open source e commerciali/no profit) e dello standard IIIF⁹. Tutto questo si realizza in un quadro di costante interlocuzione sugli standard di interoperabilità, del GIS e sulle normative di catalogazione con il Ministero della Cultura (MiC), l'ISTAT ed altri enti e coerentemente con il Piano nazionale di digitalizzazione del patrimonio culturale (PND).

BCE ha realizzato le prime applicazioni su due diocesi laziali:

1) Palestrina

La diocesi ha completato il Censimento delle Chiese, che comprende 84 edifici (83

⁷ ESRI – ArcGIS, <<https://www.esriitalia.it/prodotti/la-nostra-offerta/piattaforma-esri>>.

⁸ <https://www.beweb.chiesacattolica.it/subbeweb/>.

⁹ International Image Interoperability Framework (IIIF), <<https://iiif.io/>>.

chiese più un palazzo vescovile) descritti da una scheda CEI-A¹⁰ e 17 chiese in forma di elenco, in quanto appartenenti ad altri enti.

2) Albano

La diocesi dispone solo di un elenco di 104 edifici (102 chiese, una canonica e un palazzo vescovile).

Per entrambe le diocesi il territorio attuale rientra interamente nell'area coperta dal Catasto Gregoriano.

Il lavoro svolto ha utilizzato le mappe digitalizzate dall'Archivio di Stato e i loro *Brogliardi*, le banche dati diocesane, i *layer* cartografici CEI, il Catasto attuale, i *layer* ISTAT, la Carta Tecnica Regionale Numerica (CTRN) e la Carta Tecnica Comunale (CTC).

Il catasto cosiddetto Gregoriano è il primo catasto particellare di tutto lo Stato Pontificio, promosso da Pio VII nel 1816 e attivato da Gregorio XVI nel 1835.

L'unità di misura utilizzata per il rilievo in scala è la Canna che costituiva la misura di grandezza utilizzata prima dell'introduzione del sistema metrico decimale nello Stato Pontificio.

La mappa del Catasto Gregoriano che restituisce il rilievo eseguito sul campo è una proiezione cartografica in cui sono individuati il nord e il sud ma rispetto alla cartografia attuale gli assi sono traslati, ecco perché si rende necessaria una georeferenziazione della mappa su GIS, utilizzando punti di controllo che permettono di arrivare ad una sovrapposizione della mappa sulla cartografia attuale molto precisa.

Sono state effettuate attività di georeferenziazione delle mappe del Catasto Gregoriano: le *mappe* (1:2000), le *mappette* a scala ridotta (1:4000 o 1:8000), estrazione dai *brogliardi* (o *sommarioni*) degli edifici di culto e loro georeferenziazione, confronto in ambiente GIS tra lo stato fotografato in mappa dal Catasto Gregoriano e quanto presente oggi nelle banche dati diocesane, sul catasto attuale e sulla CTRN (1:5000) della Regione Lazio.

Per la diocesi di Albano Laziale si è presa in considerazione la sola mappa del centro urbano. Per la diocesi di Palestrina sono state georeferenziate le mappe dei centri urbani di: Palestrina, Castel San Pietro Romano, Cave, Capranica, Pisoniano, Roiate, Galliciano.

Esiti e Casi d'uso

L'utilizzo dei GIS con visualizzazione sovrapposta di più *layer* consente di confrontare in modo immediato le informazioni territoriali presenti sulle diverse banche dati (documenti di archivio, Catasto, banche dati diocesane, mappe dai geoportali nazionali o locali ecc.) e il dato storico (edificio e sua evoluzione nel tempo) per molte finalità, in particolare:

¹⁰ Progetto di schedatura degli edifici di Culto promosso da BCE della Conferenza Episcopale Italiana, <<https://beweb.chiesacattolica.it/edificidiculto/>>.

- visualizzazione completa degli immobili nel territorio, con possibile implementazione di mappe differenti sui vari *layer* (ad esempio rischio idrogeologico, classificazione sismica ecc.);
- confronto storico e toponomastico (variazione degli edifici nel tempo, denominazione edificio, demolizioni/ricostruzioni, eventi calamitosi) e con catasto attuale (ricostruzioni successive ad eventi bellici/tellurici/sventramenti urbanistici, individuabili mediante confronto stilistico-architettonico);
- analisi delle chiese dismesse e variazione d'uso;
- verifica dell'ente proprietario laddove vi sia necessità di indagine e individuazione immobili di proprietà di altri enti non censiti nelle banche dati diocesane (le *mappe* e relativi *brogliardi* del Catasto Gregoriano consentono una "fotografia" storica del patrimonio);
- verifica della georeferenziazione degli immobili, rilevante anche in caso di emergenza, e correzione anomalie;
- individuazione di complessi e chiese conventuali/confraternali (non più aperte al culto per soppressione degli enti);
- evoluzione storica di altri immobili di rilevanza a livello territoriale/di comunità.

4. Catasto Gregoriano, Censimento Chiese e GIS

Caso studio su Bologna

La CEI e l'Archivio di Stato di Roma, a luglio 2023, hanno siglato un accordo di collaborazione scientifica per l'utilizzo delle informazioni territoriali contenute nel Catasto Gregoriano nei *layer* cartografici e nelle banche dati catalografiche degli edifici di culto, realizzate dalle diocesi italiane attraverso l'applicazione della tecnologia GIS.

L'Archivio di Stato di Roma ha messo a disposizione la documentazione già digitalizzata (*mappe*, *mappette* e *brogliardi*) del Catasto Gregoriano, contenente la rappresentazione cartografica del territorio dello Stato Pontificio e i dati relativi agli intestatari delle particelle; la CEI, attraverso l'Ufficio Nazionale per i beni culturali ecclesiastici e l'edilizia di culto, sta provvedendo alla digitalizzazione di *mappe* e *brogliardi* non disponibili in tale formato e all'incrocio con i dati catalografici su supporto GIS per confrontare a livello territoriale e sociale il dato storico dei beni architettonici e le sue trasformazioni.

I risultati attesi sono nuovi servizi a beneficio delle diocesi e degli enti ecclesiastici per finalità progettuali, di gestione dell'emergenza e per la tutela, conservazione, consultazione e valorizzazione dei beni culturali ecclesiastici e un arricchimento del portale BeWeB.

Sovrapponendo le mappe opportunamente georeferenziate del Catasto Gregoriano con i dati derivati dai *brogliardi*, i dati degli edifici di culto dal software di inventariazione CEI-A¹¹, le mappe catastali attuali e la Carta Tecnica Regionale e Comunale,

¹¹ Censimento dei Beni Architettonici, <<https://bce.chiesacattolica.it/progetto-censimento-chiese/>>, <<https://beweb.chiesacattolica.it/edificidiculto/>>.

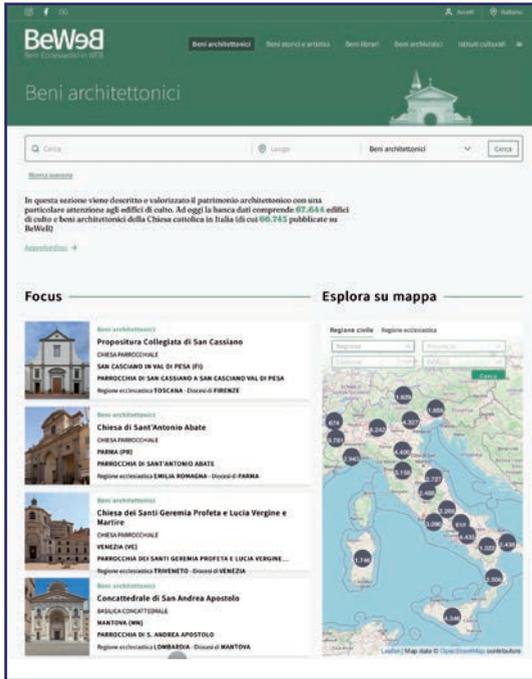


Figura 1. Il portale BeWeB, Beni architettonici

l'utilizzo dei GIS con visualizzazione su più layer (accesi/spenti) ha restituito un confronto immediato delle informazioni territoriali, consentendo:

- l'individuazione delle chiese dismesse o con variazione d'uso;
- la verifica dell'ente proprietario nei casi di immobili di proprietà di altri enti non sempre censiti nelle banche dati diocesane (il Catasto Gregoriano costituisce una fotografia storica del patrimonio al 1835).

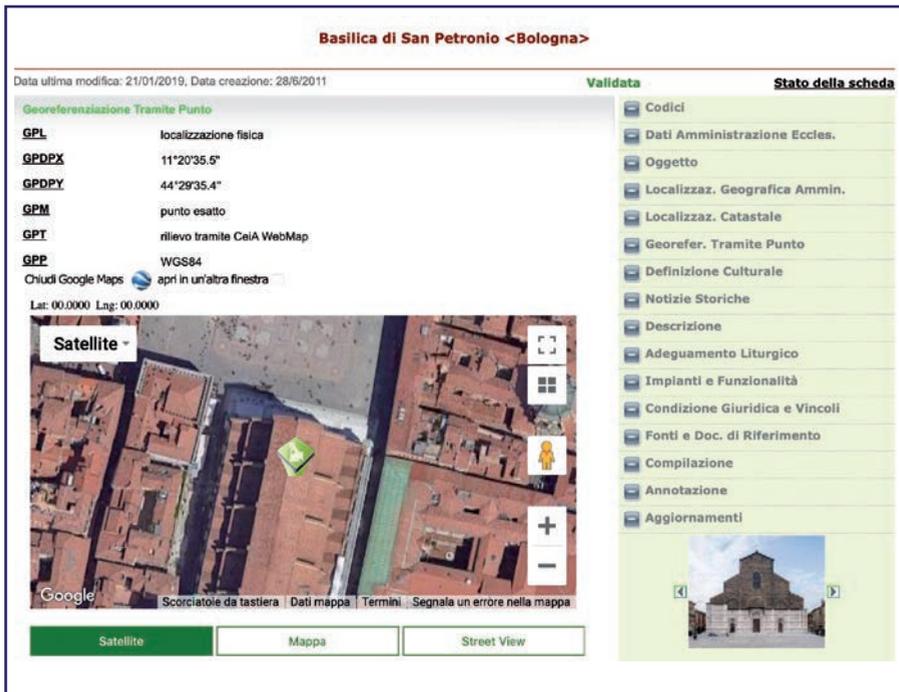


Figura 2. Scheda CEI-A



Figura 5. Dettaglio del centro storico di Bologna nel catasto attuale e nella mappa del Catasto Gregoriano

Nel caso di studio ci siamo avvalsi di alcuni dati open di un progetto di mappatura storica di fine '700 e inizio '800 del Comune di Bologna di chiese, conventi ed edifici religiosi¹².



Figura 6. Enrico Corty (disegnata, incisa), Pianta icnografica del territorio urbano di Bologna

¹² <https://www.originebologna.com>.

L'incrocio fra i *puntuali* ci ha permesso di far emergere e correggere a livello centrale alcune anomalie nella georeferenziazione degli edifici in schede catalografiche non complete o con anomalie.

Dal confronto storico e toponomastico tra i due catasti (Gregoriano e attuale) si sono rilevate demolizioni successive ad eventi bellici, tellurici o a sventramenti urbanistici conseguenti alla ri-progettazione del tessuto urbano da parte del Comune di Bologna: è il caso della Chiesa di Sant'Agata demolita in favore del Palazzo delle Poste (con contestuale modifica del toponimo, oggi "Piazza del Francia").

Sacro e profano convivevano ai piedi delle due torri: tra mercato e artigianato sotto la Garisenda sorgeva l'Oratorio della Beata Vergine delle Allegrezze, detta anche Madonna delle Grazie o della Porta, abbattuto nel 1871 per facilitare la viabilità.

È paradossale come in questi giorni per gravi problemi di staticità della torre Garisenda (a rischio crollo) sia stata interrotta precauzionalmente la viabilità di questo punto nevralgico.

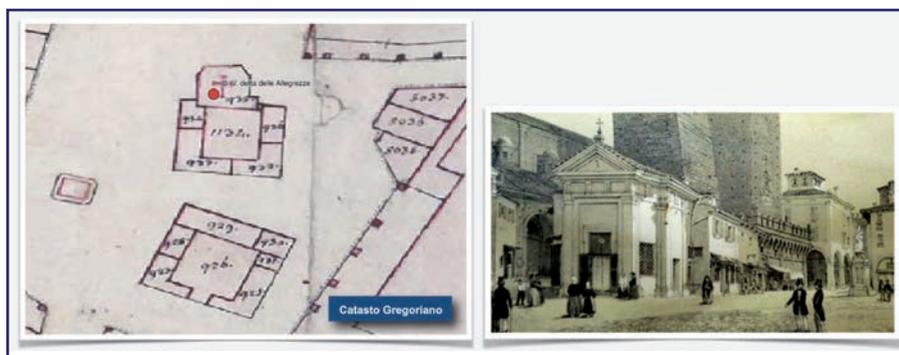


Figura 7. Chiesetta della Garisenda, abbattuta nel 1871, per facilitare la viabilità di Piazza Ravegnana. Fonte: *Le Torri di Bologna*, <<http://www.torridibologna.it>>

Il confronto costante con la banca dati di Censimento Chiese ci ha permesso di rilevare:

- ricostruzioni (in epoca diversa) sulle preesistenze, come ci confermano le notizie storiche nella scheda CEI-A;
- chiese con uso non più culturale: una biblioteca (ex Chiesa di San Giorgio, attuale Biblioteca d'Arte e di Storia di San Giorgio in Poggiale), una sede espositiva (ex Chiesa di San Mattia), un'aula magna dell'Università (ex Chiesa di Santa Lucia) e una chiesa utilizzata prima come magazzino, poi come deposito militare durante la Seconda guerra mondiale, poi come officina meccanica. Negli ultimi anni il Polo Museale dell'Emilia-Romagna ne sta curando la messa in sicurezza e il restauro (ex Chiesa di San Barbaziano);
- chiese conventuali, confraternali o di altra proprietà: in alcuni casi si tratta di chiese che rivestono un'importanza storico-artistica (ad esempio la Basilica San

Domenico) ma che non sono state inserite nel database dell'*Elenco Chiese* perché di proprietà diversa da quella diocesana.

Non si esclude che la stessa metodologia possa essere adottata anche per altri catasti storici presenti in Italia, ma da ciò che è emerso dai casi d'uso in questa prima fase di studio, ne è evidente l'utilità nell'ambito:

- della gestione degli immobili;
- di manutenzione e restauro;
- di nuova costruzione, potendo disporre della sovrapposizione di *layer* tecnici e tematici indispensabili per una corretta pianificazione dell'intervento (ad esempio rischio idrogeologico e classificazione sismica¹³).

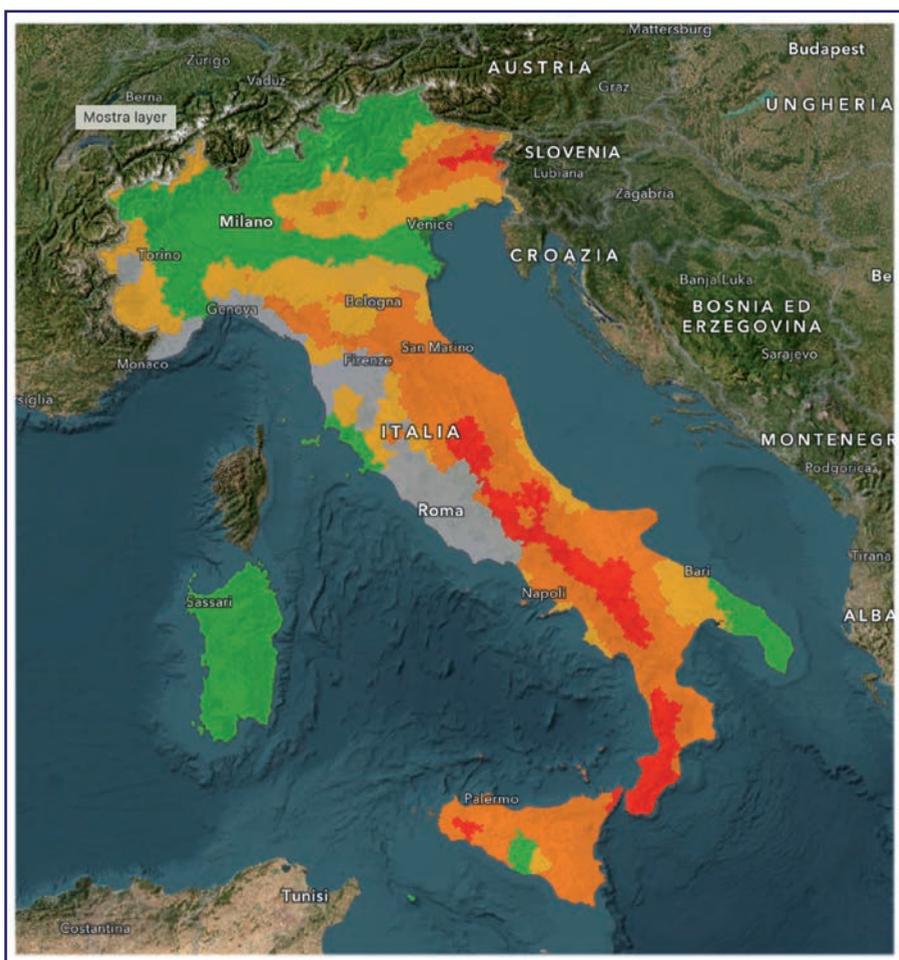


Figura 8. *Classificazione sismica 2012 dal Geoportale nazionale*

¹³ Servizi di WMS (Web Map Service) dal Geoportale nazionale, <<https://gn.mase.gov.it/>>.

The National Office for Ecclesiastical Cultural Heritage and Worship Buildings, part of the CEI, has launched a project to map the boundaries of Italian dioceses, crucial for several initiatives, such as the Church Census and Scheduled Maintenance of buildings.

Geographic Information Systems (GIS) offer an intuitive and immediate way to visualize information, allowing for the intersection of data from different sources and conducting detailed analyses.

These Systems also allow for the addition of various informational layers on a map, facilitating data understanding and analysis.

Furthermore, the utilization of information from the Gregorian Cadastre and the Church Census through GIS technologies provides new services to dioceses for managing ecclesiastical heritage, particularly for the analysis of the ecclesiastical heritage of the Diocese of Bologna.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

La prima campagna di digitalizzazione degli archivi del Venerabile Collegio Inglese di Roma

«DigItalia» 1-2024
DOI: 10.36181/digitalia-00100

Renaud Milazzo

Bibliotecario dei libri rari, Venerabile Collegio Inglese, Roma

Questo articolo ha lo scopo di descrivere la prima fase del progetto di digitalizzazione degli archivi del Venerabile Collegio Inglese di Roma. Il progetto è iniziato nel dicembre 2023 e si concluderà nel settembre 2024. Si è articolato attorno a due principali approcci: il primo mira a preservare uno dei documenti fondamentali per la storia del collegio, mentre il secondo ha l'obiettivo di rendere accessibili documenti inediti che gettano luce sulla creazione di diverse diocesi in Nord America (Québec, Baltimore e Filadelfia). La digitalizzazione del primo manoscritto ha rivelato una serie di problemi, principalmente dovuti a un restauro precedente che si è rivelato inadeguato. Questa esperienza sottolinea l'importanza di un approccio meticoloso ed esperto nella conservazione e nella digitalizzazione dei documenti storici, al fine di garantire non solo la loro sopravvivenza fisica ma anche la loro accessibilità alle future generazioni.

Il Collegio Inglese di Roma, *Venerable English College* (VEC) nella sua lingua nativa, istituzione secolare, testimonia la presenza di una comunità britannica all'interno della metropoli sin dal 1362¹. Fondato inizialmente come ostello dei pellegrini, divenne un'istituzione educativa sotto l'egida di Gregorio XIII, che promulgò la bolla papale costitutiva il primo maggio 1579. In seguito, l'istituto divenne uno strumento vitale della Controriforma, dedicato alla formazione dei missionari in divenire. Questi evangelizzatori erano incaricati di convertire le anime smarrite a vantaggio della fede riformata sotto il regno di Elisabetta I. L'istituzione giocò un ruolo preminente nell'educazione, nella preparazione spirituale e intellettuale di questi missionari, dotandoli di una profonda comprensione della dottrina cattolica e di tattiche di evangelizzazione adeguate alle sfide della persecuzione religiosa. Attraverso questo impegno, il Collegio Inglese di Roma si è imposto come baluardo della perseveranza cattolica, aspirando a ristabilire la coesione cristiana all'interno del regno d'Inghilterra e del Galles.

¹ Per una storia completa del College dalla sua fondazione si faccia riferimento a: Michael E. Williams, *The Venerable English College Rome, A History*, second edition, Herefordshire: Gracewing, 2008.

Il collegio, portatore di oltre sei secoli e mezzo di storia, conserva un corpus considerevole di documenti storici, al cui cuore si trovano atti notarili su pergamena risalenti al quattordicesimo secolo. La maggior parte della collezione è composta da registri contabili, ma comprende anche numerosi documenti architettonici relativi non solo all'edificio di Via Monserrato, ma anche a varie proprietà che sono appartenute al collegio. Essa include inoltre documenti relativi a figure eminenti, uomini e donne, legati da vicino o da lontano alla vita del seminario, quali gli Stuart, i cardinali Wiseman e Consalvi e J.R.R. Tolkien, il cui figlio fu studente al collegio. La collezione è arricchita anche da elenchi di ospiti cattolici e protestanti – tra cui John Milton, che soggiornò al seminario – e dal registro di ogni studente entrato nel seminario dal 1579, il *Liber Ruber* dal nome del colore della sua copertina (il libro rosso). A ciò si aggiunge una corrispondenza abbondante tra Roma e le diocesi di Baltimore, Philadelphia e Québec, così come migliaia di fotografie. Questi archivi costituiscono una testimonianza senza pari della vita dei cattolici inglesi e gallesi su un arco di quasi otto secoli. La singolarità di questi fondi documentari si basa tanto sulla ricchezza e l'ampiezza del loro contenuto preservato quanto sul fatto che collezioni simili di altri collegi inglesi, come quelle di Douai e di Liegi, sono state o disperse o perdute.

Inoltre, i fondi documentari di Roma si distinguono anche per la loro antichità rispetto a quelli più recenti dei collegi inglesi di Lisbona o di Valladolid. La loro sopravvivenza è in gran parte dovuta alla protezione misteriosa di un benefattore anonimo, al contrario della biblioteca del VEC che non ha conosciuto lo stesso destino. Grazie alla sua azione, gli archivi sfuggirono alle depredazioni dell'esercito napoleonico nel 1798 e furono restituiti al collegio alla sua riapertura nel 1818. Questa conservazione eccezionale ha permesso di perpetuare un'eredità storica preziosa, offrendo una prospettiva inedita sulle epoche attraversate dal collegio così come sulle dinamiche sociali, culturali, artistiche, editoriali, religiose ed economiche corrispondenti. Di fatto, questi documenti offrono una risorsa primaria essenziale per chiunque sia interessato alla storia della Chiesa, alle interazioni tra religione e politica, alle correnti di pensiero, così come agli eventi che hanno modellato l'Europa moderna e contemporanea.

Senza infrastrutture dedicate alla conservazione di questi archivi, la loro gestione è stata affidata nel 2015 all'attuale direttore, il professor Maurice Whitehead. Un primo passo è consistito nell'istituire condizioni di conservazione adeguate. Così, è stato allestito un magazzino conforme agli standard contemporanei di preservazione, facilitando l'organizzazione e la classificazione della collezione nonché l'avvio di un inventario informatizzato che risponde alle norme professionali e qualitative vigenti.

Il magazzino, oltre ad essere adeguato alle normative antincendio, è dotato anche di un sistema di controllo preciso della temperatura e dell'umidità, permettendo un monitoraggio costante e rigoroso delle condizioni ambientali.

Attualmente, è in corso una seconda fase di salvaguardia e conservazione. Il Collegio riconosce il proprio notevole ritardo nel settore della digitalizzazione e ha deciso di rimediare progressivamente. Dipendendo temporaneamente da fondi privati, i criteri per selezionare i primi documenti eleggibili per la digitalizzazione si sono basati su due fattori. Innanzitutto, sopperire al degrado fisico del *Liber Ruber*, documento fondamentale per la storia del Collegio, a causa di un restauro mal realizzato alla fine degli anni Cinquanta. In secondo luogo, mettere a disposizione della comunità scientifica un insieme di scambi epistolari che illuminano in modo unico la fondazione di diocesi oltreoceano all'inizio del diciannovesimo secolo. Per raggiungere questo obiettivo e al fine di ottenere un risultato qualitativo ottimale, la realizzazione di questo progetto è stata affidata a un fornitore esterno.

La digitalizzazione del *Liber Ruber*

Nel 1957, il *Liber Ruber*, un manoscritto in folio, ha subito un restauro che non ha sortito gli effetti sperati e che, al contrario, ha causato un'ulteriore degradazione della carta. In origine, il manoscritto soffriva di uno sbiadimento dell'inchiostro utilizzato per la scrittura, che virava dal nero al marrone rendendo difficili da leggere alcuni passaggi, e di corrosioni legate anch'esse alla composizione dell'inchiostro. Infatti, diverse lettere maiuscole non si manifestavano più a causa delle lacune della carta. Questi fenomeni sono ben noti. Il fenomeno dello sbiadimento risulta tanto dalla presenza di composti chimici residui all'interno della carta quanto dall'esposizione alla luce. L'ossidazione causa la formazione di acidi. Questa reazione chimica, congiuntamente all'aggiunta precedente di acido cloridrico o solforico nell'inchiostro per migliorarne il flusso, si rivela dannosa per la carta. Talvolta l'acidità è così invasiva da perforare la carta, trapassandola da parte a parte e raggiungendo talvolta anche i fogli adiacenti. Anche la pagina emblematica che descrive il giuramento fatto dai primi studenti del Collegio, che si impegnarono per iscritto a diventare sacerdoti e a partire per l'Inghilterra per svolgere la loro missione di evangelizzazione, rivelava righe di scrittura in via di scomparsa. Infatti, un fattore aggiuntivo ha contribuito ad accelerare la degradazione dell'opera: l'atto di devozione dei primi studenti, tra cui i futuri martiri, suscitava grande interesse tra i pellegrini e i visitatori, richiedendo così l'apertura frequente del libro a quella pagina specifica. Questa manipolazione ripetuta ha portato a un'usura significativa, indebolendo la legatura e erodendo la carta ai bordi, rendendo prima o poi necessario il ricorso a un restauro.

Nel 1957, gli specialisti del servizio degli Archivi Vaticani sono stati interpellati per la loro competenza². La prima misura presa è stata trattare alcune pagine con una soluzione bollente per estrarre il ferro dall'inchiostro, evitando così la continua deteriorazione della carta da questo metallo. Per il resto del documento, erano im-

² *Liber Ruber*, «The Venerabile, conducted by the past and present students of the Venerable English College Rome», 18 (May 1957), n. 2, p. 80-81.

maginabili due tecniche successive. La prima, un metodo moderno ancora in fase di prova, prevedeva di incapsulare il documento tra due fogli di plastica, uniti mediante pressione a caldo. Nonostante il suo basso costo, questo metodo comporta un importante svantaggio: l'incertezza riguardo alla durata del rivestimento plastico, che potrebbe diventare fragile e creparsi dopo cinquant'anni, e che, una volta applicato, è irreversibile. La seconda tecnica, provata, consiste nell'applicare sulla pagina un sottile strato di seta trasparente. La seconda opzione, considerata meno intrusiva, è stata preferita. Tuttavia, nonostante il rigoroso rispetto dei protocolli di restauro stabiliti, sembra che una vernice, probabilmente carica di una quantità significativa di plastica, sia stata incorporata durante questo processo. Sarebbe opportuno effettuare un'analisi per determinare se questa vernice sia a base di cellulosa o se sia composta da un'altra sostanza. In ogni caso, sembra che questa contribuisca non solo ad aumentare l'acidità della carta originale, che era tuttavia di altissima qualità, ma anche a provocare l'apparizione di nuove perforazioni nel documento. Normalmente, la flessibilità dei cartoni di seta impiegati nei lavori di restauro permette la loro adesione ad altri fogli mediante una colla di amido estremamente diluita (3-4%). Questo tipo di carta di seta, data la sua leggerezza, la purezza e la lunghezza delle sue fibre, risulta quasi impercettibile ed eccezionalmente resistente.

Tuttavia, il risultato del restauro del *Liber Ruber* si è rivelato diametralmente opposto a quanto atteso. Non solo lo spessore delle pagine è raddoppiato, passando da 0,12 mm a 0,24 mm, ma ciò ha anche portato a una significativa riduzione della chiarezza del documento. Di conseguenza, l'aumento dello spessore dei fogli ha reso indispensabile l'adozione di una nuova legatura; su questa coperta rinnovata, quella vecchia è stata accuratamente conservata e incorporata. Tuttavia, le preoccupazioni riguardo alla degradazione a lungo termine della vernice si sono rivelate fondate. Diventava quindi urgente avviare un protocollo di digitalizzazione del documento.

Di fronte all'assenza di attrezzature appropriate e di risorse umane necessarie per l'attuazione del progetto, il compito di digitalizzazione del manoscritto è stato affidato a un fornitore esterno, nello specifico la società Emmebissoft. Data la presenza di questa vernice sul manoscritto, che causa difficoltà aggiuntive, sono state adottate misure preventive per contrastare gli effetti nocivi sulla chiarezza e leggibilità del testo, nonché i problemi di riverbero della luce causati da questo materiale. Per superare questi ostacoli, sono state prese diverse decisioni preliminari alla digitalizzazione. Innanzitutto, riguardo alla risoluzione e definizione dell'immagine³. Piuttosto che optare per una risoluzione di 300 punti per pollice, che è me-

³ La risoluzione è un'unità di misura in punti per pollice (ppp) o *dot per inch* (dpi) che ha lo scopo di analizzare il numero di pixel contenuti in un quadrato di 2,54 cm di lato. La definizione, invece, è propria del documento e indica il numero di pixel contenuti in esso. Per ottenerla, si moltiplica il numero di pixel contenuti in altezza per quelli contenuti in larghezza.

no esigente in termini di spazio di archiviazione e che rimane conforme alla capacità di percezione visiva umana - la quale permette, a questa risoluzione, di leggere e assimilare le informazioni contenute nell'immagine digitalizzata e di giudicarla conforme all'originale -, è stato deciso di privilegiare una risoluzione di 600 dpi. Questa scelta offre la possibilità di ingrandire notevolmente i dettagli all'interno del documento digitalizzato finale. Pertanto, nelle porzioni del documento dove la vernice si rivela eccessivamente opaca, l'opzione di ingrandimento ha come obiettivo iniziale di permettere una migliore leggibilità del testo.

In secondo luogo, al fine di rimanere fedeli al manoscritto originale, è stata preferita la digitalizzazione a colori piuttosto che una digitalizzazione in scala di grigi. Di conseguenza, la profondità di codifica selezionata è di 8 bit per ciascuno dei canali di colore (rosso, verde e blu), rappresentando così un totale di 24 bit. Con l'aumento costante delle capacità di archiviazione anno dopo anno, questa decisione è sembrata corrispondere alla fedeltà richiesta dal *Liber Ruber*, poiché l'immagine ottenuta permette non solo una restituzione impeccabile su qualsiasi tipo di schermo, ma anche in caso di stampa. Infine, considerando l'insieme delle ragioni precedentemente esposte, la scelta è caduta sul formato TIFF (Tagged Image File Format) per la conservazione dei documenti. Questo formato garantisce un'alta risoluzione e una definizione superiore, assicurando così che l'immagine rimanga integrale, senza compressione né alterazione. Nonostante lo svantaggio di generare file di dimensioni considerevoli, che superano i 100 megabyte, questa decisione riafferma il nostro impegno verso l'integrità e la durabilità dei documenti digitalizzati. Offre una fedeltà visiva senza pari e assicura una vasta compatibilità con i sistemi dedicati alla gestione di archivi digitali.

Per soddisfare le esigenze di questo capitolato, la compagnia Emmebisoft, forte della sua esperienza, ha preferito costruire il proprio sistema di digitalizzazione piuttosto che affidarsi a una soluzione commercialmente disponibile. Si tratta di uno scanner dotato di un dorso digitale capace di catturare l'intero documento in un solo scatto, grazie all'uso di una o due fotocamere. Considerando la fragilità del *Liber Ruber*, che non può essere aperto completamente a 180 gradi, e richiede un'apertura compresa tra 80 e 120 gradi, l'uso di due fotocamere si è rivelato indispensabile durante il processo di digitalizzazione.

L'installazione dell'unità di produzione è stata realizzata temporaneamente all'interno del Collegio. Questo approccio ha permesso di apportare soluzioni rapide ai problemi legati alla sicurezza del trasporto del manoscritto e alle questioni assicurative. Una campagna iniziale di digitalizzazione sperimentale del manoscritto si è svolta nel dicembre 2023, risultando nella cattura digitale dell'intero insieme dei 530 fogli. Tuttavia, come previsto, la vernice ha creato un problema significativo in termini di illuminazione. Per superare questo problema, l'uso dello scanner dotato di un dorso digitale ha non solo offerto maggiore flessibilità nella scelta dell'illuminazione ma ha anche permesso di effettuare test con diverse configurazioni lumi-

nose. Dopo numerose sperimentazioni, è stato deciso di utilizzare una fonte di luce uniforme e diffusa proveniente dall'alto. Nonostante tutte queste precauzioni e un consistente lavoro di post-produzione durante il quale ogni immagine è stata sottoposta a varie correzioni, i primi risultati non sono completamente soddisfacenti.

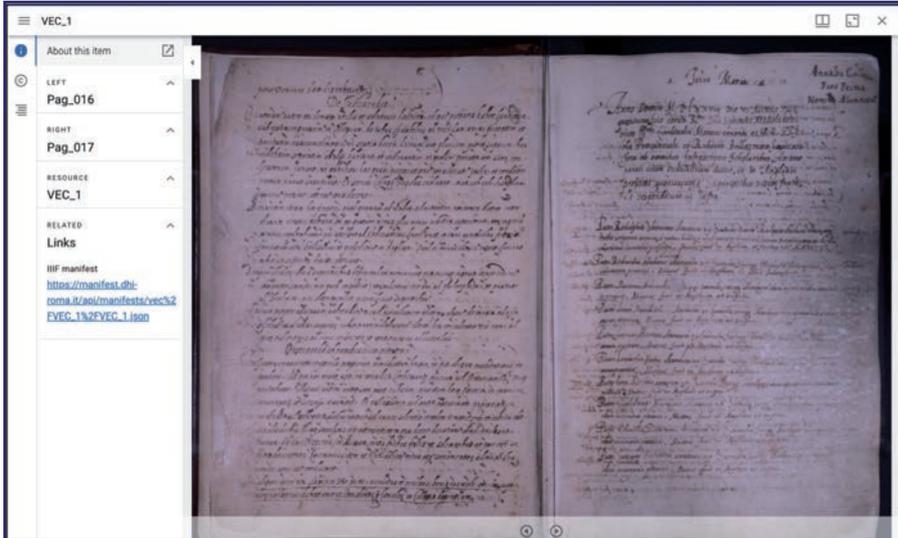


Figura 1. Esempio di risultato della prima campagna di digitalizzazione

Durante il controllo di qualità, come si può osservare nell'immagine sopra (Fig. 1), nonostante la scelta di un'alta risoluzione a 600 dpi e la registrazione in formato TIFF, la riverberazione rimane significativa. Ingrandendo l'immagine, si percepisce meglio l'impatto della pellicola di plastica sulla nitidezza della scrittura manoscritta

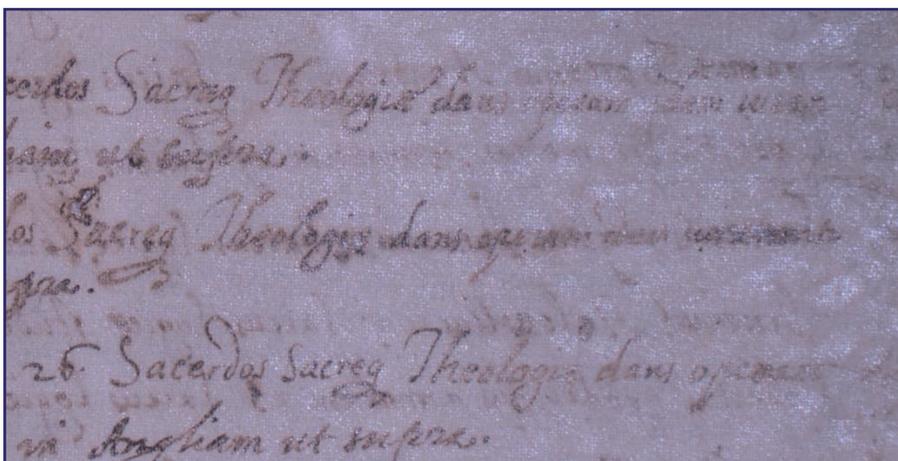


Figura 2. Ingrandendo l'immagine ottenuta appare un significativo scintillio

e la luce che crea come macchie bianche chiaramente visibili in tutta la parte destra dell'immagine e che potrebbero essere confuse con la pixelizzazione (Fig. 2). Sebbene non sia questo il caso, l'aspetto generale dell'immagine manca di contrasto e di messa a fuoco in alcune parti specifiche di ogni foglio. Di conseguenza, è stata programmata una nuova sessione di digitalizzazione per settembre 2024. In seguito a una serie di test recentemente effettuati, sembra che un sistema di illuminazione laterale, disposto ai lati del manoscritto, potrebbe ridurre notevolmente le rifrazioni luminose causate dalla vernice.

Al fine di guadagnare tempo e grazie alla prima fase di digitalizzazione, la società Emmebisoft ha iniziato il lavoro di post-produzione. In questo contesto, svolge un ruolo cruciale, con il compito di strutturare i documenti in modo coerente. Ciò implica l'uso di sistemi di gestione delle risorse digitali (Digital Asset Management, DAM) per assicurare un contesto organizzato e accessibile. Un passaggio fondamentale di questo processo è l'assegnazione meticolosa di metadati alle risorse digitalizzate. Questi metadati includono l'aggiunta di indici per una navigazione intuitiva attraverso i documenti, descrizioni dettagliate per un'indicizzazione precisa e un arricchimento semantico adeguato, didascalie esplicative, parole chiave per la ricerca e indicazioni sui diritti d'autore per rispettare la legislazione vigente. Tutte queste informazioni hanno l'obiettivo di arricchire il contesto delle risorse e ottimizzare il loro utilizzo successivo.

La digitalizzazione dei North American agency papers

Durante l'estate del 1817, Robert Gradwell (1777-1833), ecclesiastico del clero secolare britannico, ricevette l'incarico di ristabilire il Venerable English College a Roma. Infatti, il Collegio era stato requisito dalle forze francesi durante l'invasione della città eterna nel 1798 e aveva dovuto chiudere le sue porte per diverse decadi. In occasione della ripresa delle attività del VEC nel 1818, Gradwell assunse contemporaneamente e in modo officioso le funzioni di emissario per l'Inghilterra, fungendo da collegamento tra la Congregazione per l'Evangelizzazione dei Popoli e i vicari apostolici di Inghilterra e Galles. Inoltre, nel 1819, Gradwell accolse Joseph-Octave Plessis, l'arcivescovo di Québec, e lo sostenne durante il suo soggiorno romano destinato a discutere con la Congregazione per l'Evangelizzazione dei Popoli sulla divisione della sua diocesi, un argomento su cui aveva già richiesto un incontro più volte senza ricevere risposta. Successivamente, nel 1821, Gradwell accettò con riluttanza di gestire gli affari relativi al Nord America, come agente per le diocesi di Québec, Baltimore e Filadelfia. Questo ruolo fu continuato dal Dr. Nicholas Wiseman (1802-1865), successore di Gradwell come rettore del VEC dal 1828 al 1835, che in seguito fu elevato al rango di primo cardinale-arcivescovo di Westminster in occasione del ripristino della gerarchia cattolica in Inghilterra e Galles nel 1850. Di conseguenza, il Collegio conserva una serie di documenti relativi a questa funzione di Gradwell. Gli archivi dell'agenzia nord-americana sono

costituiti principalmente da scambi epistolari tra l'arcidiocesi di Baltimore e Roma e coprono un periodo che va dal 1821 al 1834. Ventitré lettere riguardano gli affari della diocesi di Filadelfia (1829-1835) e 411 fogli provenienti da diverse fonti canadesi consentono di ricostruire gli inizi della diocesi di Québec (dal 1817 al 1833). Queste risorse sono poco conosciute e richiedono di essere studiate.

In confronto con il *Liber Ruber*, la digitalizzazione dei documenti non rappresenta alcuna rilevante sfida tecnica. La digitalizzazione di oltre 500 fogli è avvenuta nel marzo 2024. Trattandosi di semplici fogli, uno scanner industriale a passaggio offre un affidabile standard qualitativo e fornisce una soluzione ottica, luminosa e di supporto integrato. È osservabile nell'immagine seguente che, al fine di semplificare il processo di catalogazione dei documenti, ogni file è dotato di riferimenti specifici al suo contenuto. Al termine delle operazioni di post-produzione, l'accesso ai documenti tramite l'interfaccia pubblica sarà facilitato dall'incorporazione, effettuata durante la catalogazione, di un link diretto.

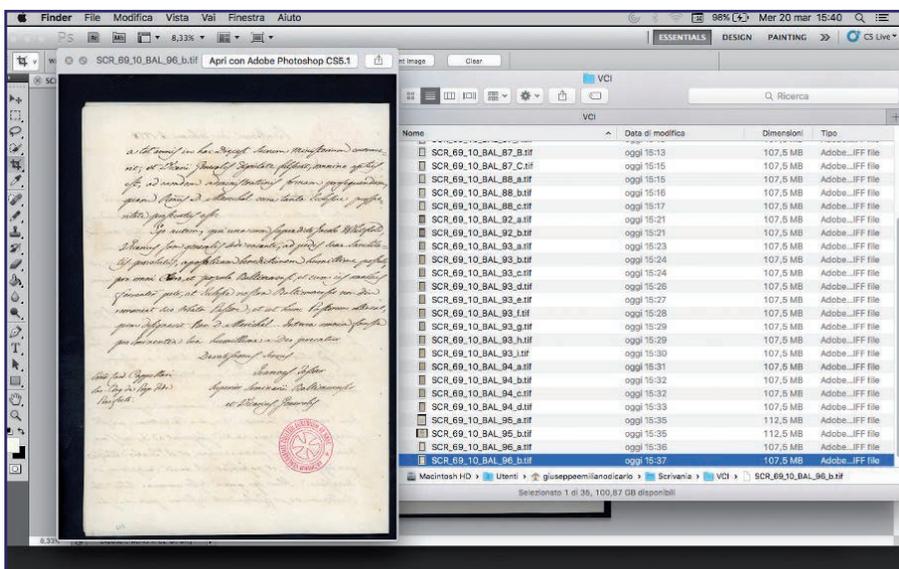


Figura 3. Esempio di organizzazione dei file digitalizzati per facilitare la catalogazione

Il lavoro di post-digitalizzazione si svolgerà nel settembre 2024 contemporaneamente al catalogo descrittivo dell'intero insieme di documenti.

Catalogazione e interfaccia pubblica

La fase iniziale di digitalizzazione avviene simultaneamente alla catalogazione dei libri antichi e rari del VEC, compito realizzato tramite il software libero Koha. Data l'alta precisione richiesta per il catalogo degli stampati, è sembrato coerente applicare standard identici alla descrizione delle diverse collezioni archivistiche.

Attualmente, non esiste un sistema perfetto per unificare le descrizioni secondo lo standard MARC21, comunemente usato per i libri, e il formato EAD, adottato per i manoscritti. Tuttavia, Koha, attraverso l'uso del linguaggio XML, offre la flessibilità necessaria per allineare le categorie descrittive utilizzate per gli archivi a quelle dei libri, garantendo così un livello uniforme di dettaglio descrittivo.

Di conseguenza, l'Opac destinato ai lettori potrà essere integrato nel catalogo collettivo URBIS, che raggruppa le principali biblioteche degli istituti internazionali a Roma⁴. Sempre in collaborazione con la società Emmebisoft, l'obiettivo strategico del VEC risiede nella centralizzazione dell'accesso alle collezioni per gli utenti, con la creazione di un portale unico. Per questi ultimi, la forma che prende la descrizione delle collezioni è secondaria; ciò che importa soprattutto è la rilevanza dei risultati ottenuti e l'eshaustività delle informazioni. Pertanto, diventa essenziale l'implementazione di un portale integrato che raccolga i cataloghi di vari formati. Questo portale dovrebbe offrire opzioni di ricerca avanzate, basate su indici unificati che saranno regolarmente arricchiti dall'incrocio delle diverse fonti disponibili. In questo modo, lettori e ricercatori troveranno raccolte in un unico luogo tutte le risorse del Venerabile Collegio Inglese.

Progetti futuri

Il progresso delle future iniziative di digitalizzazione dipenderà dalle risorse finanziarie disponibili. Uno dei progetti previsti riguarda le pergamene attualmente conservate sotto forma di rotoli, alcuni dei quali non sono stati dispiegati da secoli, il che li rende particolarmente secchi e fragili. Per questa ragione, il loro accesso in sala lettura è attualmente possibile ma strettamente controllato per minimizzare le manipolazioni che potrebbero danneggiarli. Sono già state avviate delle fasi preliminari, e la futura digitalizzazione richiederà preparativi considerevoli. La conservazione digitale potrà essere realizzata solo dopo l'appiattimento delle pergamene tramite un metodo di umidificazione delicato e non invasivo. Questo processo di trasformazione da rotolo a documento piatto presenta anche una sfida logistica significativa relativa alla conservazione e allo spazio di deposito, che dovrà essere ripensato prima della digitalizzazione.

Infine, tutti questi progetti si inseriscono in quello più ampio di creare un istituto di ricerca dedicato alla comunità cattolica di Inghilterra e Galles. Attualmente non esiste alcuna risorsa centralizzata o punto informativo in Inghilterra, Galles e all'estero per i ricercatori interessati a questo patrimonio religioso e culturale. Piuttosto che costringere le parti interessate a cercare una miriade di fonti disparate, l'istituto storico offrirà uno "sportello unico" di cui si sente fortemente la mancanza. Il successo di questo progetto non può prescindere dalla digitalizzazione e catalogazione degli archivi documentari e dei libri antichi.

⁴ Il network URBIS raggruppa un vasto insieme di biblioteche accademiche di Roma: <<https://www.urbis-libnet.org/vufind/>>.

This article aims to describe the first phase of the digitization project of the archives of the Venerable English College in Rome. The project began in December 2023 and will conclude in September 2024. It has been structured around two main approaches: the first aims to preserve one of the most fundamental documents in the College's history, while the second seeks to make accessible unpublished documents that shed light on the establishment of various dioceses in North America (Québec, Baltimore, and Philadelphia). The digitization of the first manuscript revealed a series of problems, mainly due to prior inadequate restoration efforts. This experience underscores the importance of a meticulous and expert approach in the preservation and digitization of historical documents, to ensure not only their physical survival but also their accessibility to future generations.

Documenti e discussioni

Computabilità e traduzione interlinguistica: considerazioni sull'impiego dell'Intelligenza Artificiale in ambito traduttivo

Marika Angelucci

Università "G. D'Annunzio" – Chieti-Pescara

Il presente articolo si inserisce nell'acceso dibattito attorno al settore multidisciplinare dell'Intelligenza Artificiale assumendo una postura specifica: il punto di vista traduttologico. La Traduzione Automatica, infatti, ha costituito la prima applicazione del computer in ambito umanistico, sia in virtù del generico interesse per lo studio del linguaggio umano, data la sua forte connotazione identitaria, sia per via della sua alta formalizzazione. Attraverso una panoramica delle architetture che si sono susseguite, s'intende focalizzare l'attenzione sul proficuo dialogo tra pratica e teoria, a cui si devono le riflessioni moderne sui meccanismi della traduzione, processo decisionale e di negoziazione. Tra queste: i limiti del passaggio interlinguistico, l'entropia dell'informazione, l'anisomorfismo delle lingue naturali, la mutua influenza di linguaggio e cognizione, la rilevanza delle informazioni extra-testuali e degli orizzonti cognitivi dei parlanti, il concetto di fedeltà testuale. L'approfondimento di tali aspetti apre a delle riflessioni di respiro più ampio, come la questione spinosa del rapporto tra intelligenza naturale e intelligenza artificiale, l'essenza della comprensione e il possibile impedimento euristico costituito dall'imitazione.

Introduzione

Manifestazione iper-moderna dell'aspirazione umana all'autocoscienza¹, il settore di ricerca interdisciplinare dell'Intelligenza Artificiale (IA) indaga le teorie e i metodi che consentono di comprendere, e quindi di costruire, dei modelli di razionalità, *imitativi* o meno del pensiero e dell'azione umani².

Nell'arco dell'ultimo biennio, soprattutto in seguito al rilascio della chatbot ChatGPT da parte dell'azienda californiana OpenAI, i dispositivi di IA hanno acce-

¹ Arnold Gehlen, *L'uomo nell'era della tecnica*, Roma: Armando, 2003 p. 40, in: Vallori Rasini, *L'essere umano. Percorsi di antropologia filosofica contemporanea*, Roma: Carocci, 2008, p. 117: «angustiato dall'enigma della sua esistenza e della sua stessa essenza, l'uomo non ha altra risorsa che cercare di interpretarsi passando attraverso un non io, attraverso qualcosa di diverso dall'umano».

² Peter Norvig – Stuart J Russel, *Artificial Intelligence: A Modern Approach*, Londra: Prentice Hall, 1994.

so il dibattito pubblico. Anche tra i non specialisti emergono curiosità riguardo il loro funzionamento, scetticismo o entusiasmo in merito alla possibilità di farne uso in numerosi ambiti della vita quotidiana e lavorativa e dubbi sulle politiche nazionali e internazionali necessarie a regolarne un utilizzo ragionevole e sicuro. Il presente articolo si inserisce nel dibattito attorno all'Intelligenza Artificiale, con particolare attenzione alla possibilità di sfruttare dei dispositivi per portare a termine il difficile compito della traduzione interlinguistica³.

Dalla nascita programmatica, in occasione del celebre convegno del Dartmouth College (New Hampshire) del 1956, il settore dell'IA unifica gli sforzi delle numerose discipline – la filosofia, la matematica, la psicologia, la cibernetica, l'informatica e la linguistica – che, adottando criteri e metodi specifici, hanno perseguito individualmente un obiettivo condiviso: la modellizzazione⁴ dell'intelligenza umana e, più in generale, della razionalità.

Tra le peculiarità dell'intelletto umano che ci si è proposti di ricreare attraverso degli strumenti automatici, il linguaggio, percepito come elemento fortemente identitario dell'uomo, ha da subito rivestito un ruolo chiave⁵. Proprio in virtù della sua predisposizione alla formalizzazione, non è strano che «il primo progetto di applicazione umanistica del neonato computer»⁶ sia stata proprio la Traduzione Automatica⁷: il ricorso a processi automatici al fine di trasportare i significati da un testo di partenza S , in cui sono codificati nella lingua c_s , a un testo d'arrivo T , in cui sono ricodificati nella lingua c_t .

Se il progetto di utilizzare il computer per la traduzione interlinguistica è stato il primo fra quelli concepiti al di fuori dell'ambito matematico, tuttavia, ciò è avvenuto in assenza dei presupposti teorici che, secondo la prassi tradizionale, avrebbero dovuto orientare la pratica: «l'analisi del concetto e dei metodi della traduzione sarà sollecitata proprio dagli sforzi per automatizzarla»⁸. Alla base di questo impegno, oltre agli interessi puramente intellettuali, vanno rintracciate delle importanti forze storiche: dei presupposti strategici, legati alle dinamiche di spionaggio del secondo dopoguerra, e dei presupposti economici. La possibilità di affidare la traduzione a una macchina, infatti, si riflette nell'economicità dell'attivi-

³ Si è scelto di non approfondire in questa sede le numerose e complesse questioni etiche connesse al settore dell'IA, che meriterebbero un'ampia trattazione a parte.

⁴ Tito Orlandi, *Informatica testuale. Teoria e prassi*, Roma: Editori Laterza, 2010, p. 155: «Il modello è un'entità, M , concreta o astratta, che abbia la proprietà di poter essere utilizzata per simulare e, dunque, entro certi limiti, spiegare, comprendere, il comportamento di un'entità, P , il prototipo, anch'essa concreta o astratta».

⁵ Per approfondimenti: Giosuè Baggio, *Neurolinguistics*, Cambridge: The MIT Press, 2022; Robert C. Berwick – Noam Chomsky, *Perché solo noi. Linguaggio ed evoluzione*, Torino: Bollati Boringhieri, 2016; Claudia Casadio, *Logica Informazione Automazione*, Corfino: primeVie edizioni, 2012.

⁶ T. Orlandi, *Informatica testuale. Teoria e prassi*, cit., p. 142.

⁷ Nell'articolo il termine occorre in due forme diverse: in maiuscolo quando lo si usa per indicare il settore, in minuscolo quando invece indica il prodotto della codifica.

⁸ T. Orlandi, *Informatica testuale. Teoria e prassi*, cit., p. 142.

tà traduttiva, tanto in termini di costi della prestazione, quanto di tempo di esecuzione.

Per quanto concerne il motore intellettuale della ricerca traduttologica, che in questa sede riveste il ruolo di maggiore interesse, è importante notare che, come emergerà dall'exkursus storico che segue, solo grazie alla costruzione di alcune architetture e all'osservazione dei difetti di volta in volta riscontrati nell'elaborazione dei dati è stato possibile far progredire la scienza dietro la tecnica⁹. Attraverso la sperimentazione, per tentativi ed errori, è stato possibile perfezionare la conoscenza dei meccanismi che ci si proponeva al contempo di modellizzare: ogni architettura proposta metteva in risalto nuove problematiche prima ignorate, permettendo così di approfondire la conoscenza e, conseguentemente, di mettere a punto nuovi modelli che conducessero a un risultato sempre più adeguato e soddisfacente.

L'insufficienza della traduzione diretta

La storia della Traduzione Automatica ha inizio nel 1949, quando Warren Weaver e Claude Shannon, matematici statunitensi, applicano il modello crittografico alla traduzione, da loro considerata un semplice processo di decifrazione del codice linguistico: il messaggio è codificato nella lingua di partenza e, per essere compreso dal ricevente, dev'essere individuato e tradotto, ossia rispettivamente decodificato e ricodificato nella lingua d'arrivo.

Sulla base di questa concezione, la prima architettura realizzata è stata quella "di traduzione diretta": il sistema si occupava di trovare l'equivalenza formale tra i singoli elementi della lingua di partenza e della lingua d'arrivo, sfruttando un dizionario bilingue. Certamente la traduzione *mot-à-mot* realizzata da una simile architettura non poteva essere soddisfacente: si trattava di un approccio semplicistico, fondato su una congruenza formale di fatto insussistente tra le lingue naturali¹⁰. Proprio grazie ai limiti osservati è stato possibile ripensare la pratica traduttiva e riflettere approfonditamente su questioni spinose come l'anisomorfismo linguistico, l'onnipotenza semantica e la non-linearità del codice linguistico.

Per questi motivi, l'equivalenza tra il testo di partenza e il testo d'arrivo è ottenibile a livello contenutistico, per via della capacità, condivisa dai codificatori, di razionalizzare, e quindi codificare, i mondi (soggettivamente) percepiti, non a livello formale. In altri termini, visto il principio di onnipotenza semantica, le lingue «differiscono essenzialmente per ciò che *devono* esprimere, non per ciò che *possono* esprimere»¹¹. Ogni lingua organizza in un modo che le è proprio un certo spazio di significato, reagendo all'equilibrio con la visione del mondo del parlante, e dun-

⁹ Thierry Poibeau, *Machine Translation*, Cambridge: The MIT Press, 2017.

¹⁰ Per approfondimenti: Noam Chomsky – Andrea Moro, *I segreti delle parole*, Milano: La nave di Teseo, 2022; Andrea Moro, *Le lingue impossibili*, Milano: Raffaello Cortina, 2017.

¹¹ Roman Jakobson, *Aspetti linguistici della traduzione*, in: *Teorie contemporanee della traduzione*, a cura di S. Nergaard, Torino: Bompiani, 2014 [1995].

que evidenziando certi aspetti della realtà piuttosto che altri, cogliendone alcuni elementi piuttosto che altri: le lingue naturali sono non-congruenti, non-sovrapponibili e non-lineari.

Di conseguenza, benché il tasso di fungibilità delle diverse strategie traduttive sia variabile anche in ragione della natura del testo che s'intende ricodificare, generalmente il ricorso alle "operazioni di equivalenza diretta"¹² – ossia trascrizione, prestito, calco, traduzione letterale: strategie traduttive che restano confinate alla superficie testuale – risulta inferiore rispetto alle "operazioni di equivalenza indiretta", operazioni oblique che prevedono una manipolazione informativa ben più marcata, ponendosi «all'insegna della negoziazione»¹³.

Per via dell'irriducibilità dei sistemi linguistici, la letteralità è un fenomeno raro: l'equivalenza globale tra i testi è garantita da operazioni asimmetriche – come la trasposizione, la modulazione, l'adattamento –, risultato dell'intervento plastico del traduttore. La traduzione è un "processo decisionale"¹⁴ e, sulla base della gerarchizzazione degli elementi informativi, il testo d'arrivo viene modellato attorno a un nucleo di presunta fedeltà¹⁵, con lo scopo di minimizzarne l'entropia, che pure resta ineliminabile:

«Translations are inevitably partial: meaning in a text is always overdetermined, and the information in a source text is therefore always more extensive than translation can convey. Conversely, the receptor language and culture entail obligatory features that shape the possible interpretations of the translation [...]»¹⁶.

In altri termini, T dirà *quasi* la stessa cosa di S, sia in virtù di quanto il soggetto codificatore di S concepisce ed esprime attraverso la lingua c_s , sia in virtù di quanto il soggetto codificatore di T recepisce sulla base della sua idea di mondo e ricrea attraverso la lingua c_t : i rispettivi orizzonti cognitivi costituiscono i sistemi di riferimento all'interno dei quali il testo prende forma¹⁷.

¹² Josaine Podeur, *Jeux de traduction / Giochi di traduzione*, Napoli: Liguori editore, 2008, p. 27-52.

¹³ Umberto Eco, *Dire quasi la stessa cosa*, Torino: Giunti Editore, 2003, p. 16-17.

¹⁴ Jiří Levý, *La traduzione come processo decisionale*, in: *Teorie contemporanee della traduzione*, a cura di S. Nergaard, Torino: Bompiani, 2014 [1995].

¹⁵ Umberto Eco, *Riflessioni teorico-pratiche sulla traduzione*, in: *Teorie contemporanee della traduzione*, a cura di S. Nergaard, Torino: Bompiani, 2014 [1995]; U. Eco, *Dire quasi la stessa cosa*, cit.

¹⁶ *Translation and Power*, ed. by M. Tymoczko, E. Gentzler, Amherst and Boston: University of Massachusetts Press, 2002, p. XVIII; e Susan Bassnett, *Translation*, New York: Routledge, 2014, p. 169.

¹⁷ Per approfondimenti: Antoine Berman, *La traduzione e la lettera o l'albergo nella lontananza*, Macerata: Quodlibet, 2003; Raffaella Bertazzoli, *La traduzione: teorie e metodi*, Roma: Carocci, 2019 [2006]; *Teorie contemporanee della traduzione*, a cura di S. Nergaard, Torino: Bompiani, 2014 [1995]; *Terminologia della traduzione*, a cura di M. Ulrych, Milano: Hoepli, 2002.

Le architetture di traduzione successive

Constatata l'insufficienza dell'architettura di traduzione diretta, il passo successivo verso una traduzione automatica migliore ha riguardato l'analisi preliminare del testo di partenza, filtrato e organizzato mediante l'applicazione di regole linguistiche. Nell'architettura *rule-based* emergono l'influenza delle teorie generativo-trasformazionali a essa contemporanee e la relativa presa di coscienza della struttura gerarchica del linguaggio¹⁸. Nonostante l'evidente avanzamento rispetto al passato, anche nel nuovo modello persistevano dei limiti traduttivi importanti, derivanti soprattutto dalla ristrettezza del ventaglio di informazioni di volta in volta considerate nella computazione. Il sistema, infatti, non solo continuava a lavorare su singoli elementi, ma soprattutto non aveva accesso né ai dati di co-testo, né tantomeno ai dati di contesto, irrinunciabili per la buona riuscita della traduzione, visto che nella maggior parte dei casi essi costituiscono le uniche fonti di disambiguazione strutturale e semantica.

La nascita di Internet ha rappresentato un importante punto di snodo anche nel settore traduttologico perché ha fornito ai ricercatori un enorme database non strutturato di cui servirsi. Dal confronto con la rete deriva l'architettura *corpus-based*, fondata sull'impiego dei corpora paralleli e sulla tecnica dell'*allineamento*.

Disponendo degli allineamenti di testo e traduzione, tipicamente per frase, la macchina può processare le informazioni in due modi principali: può memorizzare le equivalenze in vista delle traduzioni future e disporne in maniera cristallizzata – è il caso della *example-based translation* –, oppure può processare le corrispondenze di cui dispone per creare dei modelli statistici che orientino in modo meno prescrittivo le traduzioni future – in questo caso si parla di *Statistical Machine Translation*, SMT.

Il secondo caso, ossia l'applicazione dei metodi statistici alla traduzione, ha segnato il vero punto di svolta della Traduzione Automatica perché ha prodotto la cosiddetta "equazione fondamentale", madre di tutti i modelli traduttivi successivi: $T = \operatorname{argmax}(T) [\operatorname{Pr}(S|T) * \operatorname{Pr}(T)]$ ¹⁹.

Questa formula matematica condensa in pochi segni tutto lo stato dell'arte: come simboleggiato dalla funzione *argmax* – che esprime la ricerca dell'elemento del Dominio della funzione la cui immagine è massima – tradurre, in termini di computazione, significa trovare, dato un testo di partenza *S*, il migliore – cioè più probabile – testo d'arrivo *T*. Nel dettaglio, la probabilità del testo di partenza dato il testo di arrivo, $\operatorname{Pr}(S|T)$, rappresenta il "modello traduttivo" (*translation model*), un sistema che conosce le possibili traduzioni di unità della lingua *S* in lingua *T* e che, sfruttando dei corpora paralleli, misura la probabilità che la sequenza di parole *S* corrisponda alla sequenza di parole *T*; la probabilità di *T*, $\operatorname{Pr}(T)$, rappresenta il "mo-

¹⁸ Per approfondimenti: Noam Chomsky, *Il linguaggio e la mente*, Torino: Bollati Boringhieri, 2010 [2006]; Steven Pinker, *Words and rules*, New York: Basic Books, 2015 [1999].

¹⁹ Thierry Poibeau, *Machine Translation*, Cambridge: The MIT Press, 2017, p. 126-131.

dello linguistico” (*language model*), un sistema che sfrutta dei corpora monolingui e misura la probabilità che T sia una sequenza di parole valida in base alle regole proprie della lingua d’arrivo. L’equazione fondamentale ha, quindi, complessivamente il significato seguente: la traduzione migliore (T’) è quella per cui il “modello traduttivo” e il “modello linguistico” risultano ottimizzati.

Come si nota dall’unità tipica di allineamento, a differenza delle architetture precedenti la SMT consente di lavorare alla traduzione gestendo le informazioni relative alla frase intera e non più solo a singoli elementi. Tuttavia, anche in questa tipologia di architettura persistono delle debolezze legate tanto alla natura dei corpora che, per quanto estesi, sono strutturalmente lontani dall’esaustività delle combinazioni linguistiche, quanto alla perenne inaccessibilità al contesto e al co-testo, se non di immediata adiacenza al segmento testuale considerato.

Conseguentemente, il passo successivo è stato mosso proprio verso una migliore integrazione nel modello delle informazioni co-testuali e contestuali necessarie alla decodifica del significato del testo di partenza e, quindi, alla formulazione di un output che rispettasse i criteri di testualità, primo fra tutti la coerenza.

Questa, per definizione, non è rintracciabile nella forma materiale del testo, non è connessa al testo nella sua dimensione concreta, ma è il risultato di un processo, dell’interazione tra testo e soggetto: il soggetto interpretante, ponendosi in maniera attiva di fronte al testo, ne inferisce il significato facendolo reagire con le informazioni extratestuali in suo possesso, stagliandolo sul suo *background knowledge*²⁰.

Affinché la coerenza venga rispettata, si rende necessario integrare nel modello le informazioni extratestuali utili affinché il testo possa essere codificato e ricodificato correttamente. Viste la plasticità e la dinamicità delle idee e del sapere, sarebbe stato poco funzionale cercare la soluzione nella creazione di un database: per quanto esteso, la sua staticità lo avrebbe reso di rapida obsolescenza; pertanto, si è pensato di focalizzare l’attenzione non sulla quantità dei contenuti accessibili, ma sul sostrato e sui meccanismi grazie ai quali tali contenuti vengono generati nella mente umana.

Le reti neurali

Sono questi i presupposti della *Neural Machine Translation* (NMT), basata sull’architettura della “rete neurale” e sul meccanismo fondamentale dell’*apprendimento*. Le reti neurali sono dei sistemi computazionali complessi d’impianto connessionista, concepiti sin dagli albori del settore dell’Intelligenza Artificiale, ma riscoperti e utilizzati sistematicamente solo a partire dal 2010, momento in cui il resto dell’apparato tecnologico ha potuto soddisfare la richiesta di una simile potenza di calco-

²⁰ Maria E. Conte, *Condizioni di coerenza. Ricerche di linguistica testuale*, Alessandria: Edizioni dell’Orso, 1999.

lo. Il modello della rete neurale si ispira alla struttura fisiologica comunemente ritenuta essere il sostrato dell'intelligenza naturale: il *connettoma*, l'insieme delle connessioni tra i neuroni del sistema nervoso, la «mappa di navigazione del nostro cervello»²¹. Composta da nodi e vettori, la rete neurale modella i pirenofori e gli assoni biologici: i nodi rappresentano dei concetti e i vettori i mezzi di trasmissione dell'impulso. A ogni vettore è associato un peso (*weight*) che esprime quantitativamente l'efficacia sinaptica della connessione tra i nodi interessati: più un vettore è "pesante", più quella connessione è forte e, quindi, probabile e stabile.

Il percorso neurale più idoneo all'elaborazione dell'informazione viene stabilizzato grazie al processo di *apprendimento*: poiché è improbabile che le stime iniziali della macchina siano soddisfacenti, è necessario che questa venga allenata (nella fase di *training*) a processare le informazioni secondo le aspettative dell'utilizzatore; per ottenere dei risultati conformi, alla macchina vengono sottoposti numerosi esempi da elaborare, su cui l'operatore interviene con delle correzioni. Sulla base di tale *editing*, la macchina ripercorre a ritroso il percorso individuato dai nodi, rivede il peso dei vettori e stabilisce il cablaggio che conduce all'output migliore, ossia all'ottimizzazione dei valori numerici.

Come nel caso dell'architettura *corpus-based*, la *Neural Machine Translation* fa ricorso ai corpora paralleli e all'analisi statistica; tuttavia, diversamente da quanto avveniva con le architetture precedenti, la macchina, a partire da queste stesse risorse, non si limita più a eseguire un'associazione lineare, ma riesce a sfruttare in maniera dinamica le informazioni di cui dispone, creandone una mappa. In effetti, siccome la rappresentazione del testo è affidata a interi gruppi di nodi-concetto, è distribuita su interi strati (*layers*) di unità informative, diventa possibile creare uno schema analitico dei significati e, conseguentemente, veicolare il testo dalla lingua di partenza alla lingua d'arrivo senza quei vincoli formali che tanto costringevano le architetture precedenti.

Il panorama attuale dei dispositivi di IA pensati per il trattamento del linguaggio verbale è dominato dai Large Language Model (LLM), architetture neurali estremamente ampie di tipo generativo: dato un *prompt*, la rete genera automaticamente, in maniera non supervisionata, delle sequenze di parole, concatenate sulla base dei dati statistici acquisiti nella fase di *training*. Se implementate in software per la traduzione, queste architetture ne rappresenterebbero uno strumento di predizione statistica senza pari.

Un esempio tra tutti, GPT-3, la terza versione del *Generative Pretrained Transformer* dell'azienda californiana OpenAI: una rete neurale formata da novantotto *hidden layers*, costruiti su centosettantacinque miliardi di parametri. Si tratta di un modello di linguaggio autoregressivo – la variabile in uscita dipende linear-

²¹ Sebastian Seung, *Connettoma, La nuova geografia della mente*, Torino: Codice Edizioni, 2016, p. 8.

mente dai valori delle uscite precedenti – che utilizza il meccanismo dell’apprendimento per generare automaticamente delle sequenze di parole a partire da un input sorgente²².

Della stessa azienda si ricordano anche ChatGPT – la chatbot che ha suscitato l’interesse generale, anche dei non specialisti, accendendo di recente il dibattito pubblico sui temi dell’Intelligenza Artificiale e della privacy dei dati²³ – e GPT-4, ampio modello multimodale (da immagine o testo a testo) rilasciato a marzo 2023, che costituisce l’evoluzione *scaled* di GPT-3.

Potenzialità e limiti della traduzione automatica

Il ricorso alla Traduzione Automatica offre numerosi vantaggi, tra cui:

- la convenienza economica, tanto in termini di costi della prestazione, quanto in termini di tempo di realizzazione dell’elaborato; da cui consegue...
- la democraticità dell’accesso ai contenuti codificati nel testo: tradurre rapidamente e a basso costo significa garantire a un pubblico ampio l’accesso a contenuti altrimenti non facilmente fruibili;
- la numerosità delle combinazioni linguistiche a disposizione: la quantità delle combinazioni linguistiche, attive e passive, coperte da un singolo software, e quindi la quantità di testi che un singolo software è in grado di gestire ed elaborare, supera di gran lunga le potenziali capacità di un traduttore umano.

Se i vantaggi appaiono chiari e ben determinati, il profilo dei limiti della Traduzione Automatica si delinea meno nettamente, perché la loro valutazione implica delle riflessioni e delle argomentazioni plurivoche.

Al di là delle problematiche proprie della traduzione *tout court*, legate alle dinamiche della negoziazione dei significati e della dispersione entropica dell’informazione, la macchina deve misurarsi con ulteriori difficoltà, che derivano dal confronto tra la sua struttura costitutiva, che ne determina il funzionamento, e il linguaggio umano, tra cui:

²² Per approfondimenti: Fabio Ciotti, *I Large Language Model e i “problemi difficili”*, «Informatica Umanistica e Cultura Digitale: il blog dell’ AIUCD», 22 agosto 2023; Luciano Floridi – Massimo Chiriatti, *GPT-3: Its Nature, Scope, Limits, and Consequences*, «Minds and Machines», 30 (2020), p. 681-694; Katherine Elkins – Jon Chun, *Can GPT-3 Pass a Writer’s Turing Test?*, «Journal of Cultural Analytics», 5 (2020), n. 2, <<https://doi.org/10.22148/001c.17212>>; Will D. Heaven, *OpenAI’s new language generator GPT-3 is shockingly good—and completely mindless*, «MIT Technology Review», 20 luglio 2020; Gary Marcus, *The New Science of AI Intelligence*, The Road to AI We Can Trust, 14 maggio 2022; Justin Weinberg, *Philosophers On GPT-3 (updated with replies by GPT-3)*, «Daily Nous», 30 luglio 2020.

²³ Per approfondimenti: Greg Brockman, *The Inside Story of ChatGPT’s Astonishing Potential*, TED Talk, 20 aprile 2023; Gabriel A. Silva, *The ChatGPT Debate: Are We Intelligent Enough To Understand ‘Intelligence’?*, «Forbes», 14 marzo 2023; Noam Chomsky, *The False Promise of ChatGPT*, «The New York Times», 8 marzo 2023; Riccardo Staglianò, *ChatGPT, meraviglie e pericoli dell’intelligenza artificiale*, «Repubblica», 19 aprile 2023; Alessia Conzonato, *ChatGPT è diventata “più stupida”? Le teorie degli utenti sulla regressione*, «Corriere della Sera», 30 luglio 2023.

- l'imprevedibilità data dalla creatività linguistica, una caratteristica peculiare e pervasiva del linguaggio umano diametralmente opposta ai requisiti di esplicitazione e di alta formalizzazione necessari al corretto funzionamento della macchina;
- la mancanza di senso comune, che mina la possibilità stessa di rispettare la coerenza testuale.

In generale, poiché tradurre significa ricercare l'identità di senso e di registro tra un testo di partenza e un testo di arrivo, con il passaggio tanto dei semi inerenti, quanto dei semi afferenti, nonché, nei limiti del possibile, della parte di senso veicolata dalla forma, la creatività linguistica, nel caso della traduzione, diventa uno strumento irrinunciabile per la gestione del livello di entropia. La necessità di svincolarsi dalla forma del testo di partenza per veicolare l'insieme dei significati in un testo d'arrivo che si inserisca correttamente nella prospettiva cognitiva e culturale del ricevente fa della creatività linguistica un requisito indispensabile per la buona riuscita della traduzione.

Concentrando l'attenzione sui fenomeni individuali, un caso particolarmente significativo è rappresentato dalle espressioni metaforiche, figure del linguaggio estremamente complesse e altrettanto frequenti, la cui peculiarità è l'inconsistenza semantica: creano un ponte tra due ambiti semantici separati, anche molto distanti, che condividono un solo seme di significato. Ancor prima che retorico, il ruolo rivestito dalle espressioni metaforiche è cognitivo: sono strategie distintive del pensiero umano che svolgono una funzione euristica. Elena Semino, linguista angloitaliana, spiega, infatti, che la figura retorica concreta non è altro che la materializzazione linguistica di un meccanismo mentale, detto "metafora concettuale", che permette di spiegare qualcosa di nuovo e astratto attraverso il concreto, ampiamente noto²⁴. Se la metafora prevede un salto tra campi semantici, tanto più lungo quanto più l'espressione è ardita, evocativa ed efficace, allora, vista la sua scarsa predicibilità, rappresenta una vera sfida per la formalizzazione. Al contempo, però, il problema della codifica dello scarsamente prevedibile si pone con urgenza: essendo uno strumento cognitivo irrinunciabile, "di cui viviamo"²⁵, la metafora si manifesta linguisticamente molto di frequente e in ogni tipologia testuale.

Accanto al limite della creatività, un altro ostacolo materiale alla traduzione automatica ottimale è la mancanza di accesso al senso comune. Il *commonsense reasoning* è un altro esempio di meccanismo primario della mente umana, che impara a memorizzare gli andamenti, le relazioni di causa-effetto, i fenomeni tipici della realtà che abita e a tenere in considerazione i dati così inferiti nelle proprie valutazioni. Senza la conoscenza di come funziona il mondo e delle situazioni che vi hanno tipicamente luogo per un corpo che lo abita, spesso diventa difficile rispettare il criterio fondamentale di testualità, ovvero la coerenza.

²⁴ Elena Semino, *Metaphor in Discourse*, Cambridge: Cambridge University Press, 2008.

²⁵ George Lakoff – Mark Johnson, *Metaphors We Live By*, Chicago: University of Chicago Press, 2003.

Proprio con la volontà di colmare questa lacuna, Hector Levesque, docente di Computer Science presso l'Università di Toronto, ha proposto di valutare l'intelligenza della macchina attraverso un quesito a scelta binaria la cui soluzione richiede l'accesso a informazioni extratestuali che l'essere umano considera del tutto ovvie, ma che tali non sono per un calcolatore, per cui tutto dev'essere esplicito e correttamente formalizzato²⁶. Il cosiddetto "Schema di Winograd" è costituito, quindi, da una domanda tripartita, così strutturata:

- una frase che contenga due sintagmi nominali e un pronome ambiguo, che potrebbe formalmente riferirsi a entrambi i sintagmi;
- una domanda sul sintagma a cui si riferisce il pronome;
- due risposte relative ai sintagmi nominali.

Apparentemente, si tratterebbe della semplice risoluzione di un'anafora; in realtà la conoscenza di fondo e il senso comune sono strumenti indispensabili all'individuazione della soluzione esatta.

«The trophy would not fit in the brown suitcase because it was too big. What was too big?

Answer 0: the trophy.

Answer 1: the suitcase».

Formalmente, il pronome neutro contenuto nella prima frase potrebbe riferirsi tanto a *trophy* quanto a *suitcase*; appare evidente, però, che, per via dei meccanismi che caratterizzano il funzionamento della realtà, solo una è la risposta plausibile: il pronome *it* non può che sostituire *trophy*, perché un trofeo troppo grande potrebbe non entrare nei volumi di una valigetta, laddove una valigetta troppo grande potrebbe contenerlo fin troppo.

«Since it was raining, I carried the newspaper over my backpack to keep it dry. What was I trying to keep dry?

Answer 0: the backpack.

Answer 1: the newspaper».

Anche in questo caso, il secondo pronome *it* potrebbe ipoteticamente riferirsi a entrambi i sintagmi nominali, ma l'evocazione della scena è dirimente: la soluzione non può che essere la prima.

Tuttavia, in entrambi i casi, senza l'accesso al senso comune e limitandosi alla superficie linguistica, le risposte 0 e 1 risultano egualmente valide: il riconoscimento dell'erroneità di una delle due risposte non dipende da questioni formali, bensì da parametri extralinguistici, di verosimiglianza, di rispondenza alle dinamiche che caratterizzano la realtà.

²⁶ Hector Levesque, *The Winograd Schema Challenge*, Toronto: University of Toronto, 2011.

Il problema della comprensione

Il problema precedentemente sollevato in termini di senso comune consente di riflettere sulla questione più ampia della comprensione che, nonostante la diffusa fascinazione – di certo alimentata dal lessico opaco, personificante e ingannevole ampiamente adottato nel mondo dell’Intelligenza Artificiale –, non sembra rientrare nelle capacità della macchina.

La meccanicità dei dispositivi di IA, concepiti su basi informatizzate e computabili, è celata da un velo di mitizzazione, infittito da «mesmerici effetti antropomorfici»²⁷ che sono indizio del successo della simulazione. Ma l’identità dell’effetto non implica l’identità del sostrato che ne è la causa: il computer, anche quando dotato di una capacità di calcolo difficilmente concepibile per la mente umana, continua a lavorare in un modo rigorosamente computabile, molto diverso da quanto riconosciuto nella dinamica della mente dalle teorie più recenti; la complessità viene erroneamente scambiata per trascendenza.

Il computer «è solo volgare silicio [...] sono solo elettroni che vanno e vengono»²⁸. Affermazioni simili sottintendono l’incommensurabilità tra uomo e macchina, una differenza ontologica giustificabile in vario modo. Limitatamente alla conoscenza attuale, l’impossibilità del confronto è motivata dalla diversa strutturazione della materia, che ha conseguenze sulla natura delle interazioni tra gli elementi: nel caso della macchina, l’interazione delle unità di materia sembrerebbe trovare piena spiegazione con la sola fisica classica, del tutto computabile; nel caso dell’uomo, né la fisica classica né la fisica quantistica tradizionale – tra l’altro, non computabile per definizione – sembrerebbero bastare e, secondo Roger Penrose, matematico britannico Nobel per la Fisica, bisognerebbe riferirsi a una fisica quantistica anomala per illustrare gran parte della fenomenologia della mente²⁹.

Certamente sono innumerevoli i quesiti ancora aperti nell’ambito dello studio della mente e l’indagine scientifico-filosofica tenta di approssimarsi alle risposte attraverso l’adozione di prospettive teoriche diverse, sia in termini di multidisciplinarietà dei contributi, sia in termini di varietà delle concettualizzazioni della mente di riferimento – dagli approcci ormai minoritari, come quelli relativi a forme di dualismo o di riduzionismo estremo, fino al funzionalismo computazionale imperante, un fysicalismo non riduzionista³⁰.

All’interno del dibattito, la questione della comprensione sarebbe focale in quanto componente funzionale della stessa intelligenza naturale: è il processo elaborativo, complesso e prolungato, attraverso cui razionalizziamo il mondo che ci circonda; è

²⁷ Henry Shevlin, *A Digital Remix of Humanity*, in: Justin Weinberg, *Philosophers On GPT-3 (updated with replies by GPT-3)*, «Daily Nous», 30 luglio 2020.

²⁸ Tito Orlandi, *Informatica testuale. Teoria e prassi*, cit., p. 164.

²⁹ Roger Penrose, *The Emperor’s New Mind*, Oxford: Oxford University Press, 2016 [1989].

³⁰ Michele Di Francesco — Massimo Marraffa — Alfredo Tomasetta, *Filosofia della mente. Corpo, coscienza, pensiero*, Roma: Carocci, 2023 [2017].

lo strumento grazie al quale possiamo spostarci dal piano del presente, dell'analisi e dell'azione sincroniche, al piano del futuro, della pianificazione e della previsione; è il mezzo che permette di *generalizzare*, ossia di astrarre delle regolarità dietro le diverse apparenze fenomenologiche.

Sebbene le posizioni a riguardo siano molteplici, in generale la comprensione si delinea come un procedimento olistico che non emerge dal rispetto di regole algoritmiche.

Secondo Tito Orlandi – in analogia con la caratterizzazione dell'inconscio fornita dagli studi di scienza cognitiva³¹ e in accordo con la concezione fodoriana della *sopravvenienza* del mentale sul fisico³² – la comprensione sarebbe il risultato della combinazione di due tipologie di conoscenza, ambedue radicate nel soggetto che conosce, non nella cosa conosciuta:

- la conoscenza noetica, che è intrapsichica, non strutturata e, di conseguenza, non comunicabile;
- la conoscenza comunicabile, strutturata, ossia l'informazione³³.

Dalle due definizioni di Orlandi si inferisce che l'unica conoscenza ontologicamente accessibile al computer è la seconda; una simile restrizione fa sì che l'informazione cui può accedere un calcolatore risulti essere parziale, incompleta rispetto a quanto decodificato dalla mente umana. Se solo la combinazione di una conoscenza formalizzabile, e quindi comunicabile, e una conoscenza intuitiva, non formalizzabile e quindi non comunicabile, può risultare nella comprensione, allora questa non può che restare preclusa alla macchina, per motivi costitutivi, poiché la macchina non ha i mezzi per elaborarla.

Se anche nell'ambito dell'Intelligenza Artificiale si parla ambigualmente di *apprendimento*, è bene sottolineare che quello della macchina è di tipo meramente meccanico, procedurale, matematico: non si svincola in nessun modo dai meccanismi materiali di computazione e non implica in nessun modo la comprensione. È una strategia algoritmica di miglioramento dei dati numerici, in particolare di riduzione della "funzione di errore", il cui modulo è inversamente proporzionale alla qualità del risultato dell'elaborazione.

Per via della natura del suo funzionamento e indipendentemente dalla complessità e dalla numerosità dei calcoli che è in grado di eseguire, la macchina non accede alla comprensione; una simile limitazione ha delle dirette conseguenze nei termini della decodifica e della ricodifica testuali.

³¹ Massimo Marraffa — Alfredo Paternoster, *Sentirsi esistere. Inconscio, coscienza, autocoscienza*, Roma: Editori Laterza, 2013, p. 21: «i processi subpersonali della scienza cognitiva hanno caratteri diversi dalla coscienza: dove questa si presenta come unitaria, seriale, linguistica e sensibile a proprietà globali, quelli sono molteplici, paralleli, non linguistici e orientati all'elaborazione di proprietà locali».

³² M. Di Francesco — M. Marraffa, A. Tomasetta, *Filosofia della mente. Corpo, coscienza, pensiero*, cit.

³³ T. Orlandi, *Informatica testuale. Teoria e prassi*, cit., p. 146-147.

In effetti, il computer è una mera macchina semiotica, un manipolatore di veicoli segnici: solo il soggetto umano, attraverso la sua competenza, caricandolo di significato, è in grado di trasformare il veicolo segnico in segno³⁴. In altri termini, la conoscenza non è materialmente connessa ai dati codificati nella macchina, ma risiede nella mente, che reagisce attivamente allo stimolo del veicolo segnico, facendosi protagonista di un processo di decodifica-codifica. Per dirlo ancora con le parole di Tito Orlandi:

«il significato in senso proprio dei veicoli segnici gestiti dai computer, cioè i gruppi di bit, è attivo soltanto in presenza dei soggetti umani, mentre per i computer essi portano con sé un significato in senso molto particolare, cioè significano semplicemente se stessi, sono riconosciuti in quanto uno è diverso dagli altri [...]. La competenza umana rinvia invece al significato dei segni in quanto unità dell'informazione, diciamo così, esterna, cioè non ciò che essi sono, ma ciò che essi rappresentano»³⁵.

Attraverso il glifo, l'occhio umano legge una rete di sensi, astrae le immagini e i contenuti evocati dal veicolo segnico materiale, che sta per qualcos'altro. Al contrario, per la macchina il glifo rappresenta solo sé stesso, un ente materiale che si combina con altri, innescando dei meccanismi di interazione e sostituzione con gli altri glifi come fossero pedine su una scacchiera.

La visione di Orlandi è coerente con quanto sostenuto da Searle in merito al rapporto tra sintassi e semantica e al confronto tra mente e Macchina di Turing: la semantica non è riducibile alla sintassi e dunque le operazioni mentali non possono essere definite in termini di manipolazioni meramente sintattiche³⁶.

Dal punto di vista traduttologico, adottando questa cornice teorica di riferimento³⁷, è evidente che la limitatezza della parte formalizzabile dell'apparato conoscitivo connesso al testo si riflette nell'incompletezza delle informazioni computabili, e dunque veicolabili, dalla macchina. Elementi di conoscenza noetica saranno sempre presenti, sebbene in tasso variabile e in modo più o meno evidente, anche in ragione della tipologia testuale in analisi: testi di natura diversa saranno, di conseguenza, più o meno adatti a essere sottoposti al trattamento automatico, e richiederanno un *editing* umano più o meno marcato, che possa colmare o comunque gestire al meglio le lacune informative. Ad esempio, il rapporto di biunivocità tra significante e significato che caratterizza gran parte del lessico utilizzato in un testo specialistico o in un testo pratico non è generalmente rintracciabile in un testo

³⁴ Ivi, p. 152.

³⁵ Ivi, p. 148.

³⁶ John R. Searle, voce *Chinese Room Argument*, in Robert A. Wilson — Frank C. Keil, *The MIT Encyclopedia of the Cognitive Sciences* (MITECS), Cambridge, Mass.: MIT Press, 1999, p. 115-116.

³⁷ Per approfondimenti: Dino Buzzetti, *Turing e l'elaborazione automatica del testo*, in: Accademia Nazionale dei Lincei, *Per il centenario di Alan Turing fondatore dell'informatica*, Roma: Scienze e Lettere, 2014, p. 159-188.

letterario o in un testo giornalistico, tipologie testuali che sfruttano l'ambiguità, in cui si ricorre ampiamente a giochi sulla polisemia o sulla forma delle parole.

Una strada sbagliata?

Se le teorie sulle anomalie quantomeccaniche della mente di Penrose e sull'irriducibilità della semantica alla sintassi di Searle fossero corrette, allora le neuroscienze e il settore dell'Intelligenza Artificiale sarebbero attualmente in un vicolo cieco: nella prospettiva della modellizzazione dell'intelligenza generale, la riproduzione del connettoma biologico mediante le reti neurali sarebbe insufficiente, perché ne tralascerebbe una parte costitutiva essenziale; ma l'inclusione di tale componente fondamentale richiederebbe un kuhniano cambiamento di paradigma, che consentisse il superamento dei limiti imposti dalla computabilità³⁸.

Focalizzando l'attenzione sulle basi ideologiche rintracciabili nell'approccio attuale allo studio di quanto concerne la mente umana e alla progettazione dei dispositivi di IA, sembra plausibile ricondurre l'insufficienza dei risultati finora conseguiti a due *bias* fondamentali, due preconcetti latenti fuorvianti che andrebbero problematizzati e contestati:

- che l'intelligenza umana sia del tutto, o quantomeno in massima parte, inquadabile in termini di razionalità, secondo la «secolare e imperturbabile formula dell'uomo "animale razionale"»³⁹;
- che l'intelligenza umana rappresenti una forma di pensiero gerarchicamente superiore alle altre possibili.

A questo proposito, alcuni esponenti della comunità scientifica si mostrano scettici riguardo l'utilità dell'approccio imitativo in generale. Secondo Rodney Brooks, professore emerito di robotica presso il Massachusetts Institute of Technology (MIT), per esempio, il limite delle neuroscienze e del settore IA, ancor prima che pratico – il che sarebbe tanto più vero nel caso in cui la teoria di Penrose non venisse falsificata –, sarebbe concettuale e in qualche modo autoimposto: gli intenti imitativi adottati da gran parte della ricerca sarebbero controproducenti, perché, oltre a essere un riflesso del narcisismo umano, ingabbierebbero la mente dei ricercatori, adeguandola a uno schema di pensiero prestabilito. In questo modo sarebbe loro preclusa la libertà di concepire dei modelli di razionalità del tutto nuovi, svincolati dai meccanismi della tanto venerata mente umana. «*We are in an intellectual cul-de-sac, in which we model brains and computers on each other, and so prevent ourselves from having deep insights that would come with new models*»⁴⁰.

³⁸ Per approfondimenti: Will Knight, *One of the fathers of AI is worried about its future*, «MIT Technology Review», 17 novembre 2018; Gary Marcus, *The deepest problem with deep learning*, «Medium», 1 dicembre 2018.

³⁹ Vallori Rasini, *L'essere umano. Percorsi di antropologia filosofica contemporanea*, Roma: Carocci, 2008, p. 120.

⁴⁰ Rodney Brooks, *Avoid the cerebral blind alley*, in: *Turing centenary. Is the brain a good model for machine intelligence?* «Nature» (22 febbraio 2012), n. 482, p. 462-463, <<https://doi.org/10.1038/482462a>>.

Come riporta Johannes Bruder, direttore dell'Istituto di ricerca in Design Sperimentale e Media Culture (IXDM) di Basilea:

«While the analogy between human brain and computer might have been fruitful for both computing and neuroscience over the course of the 20th century, the belief that computers could become “intelligent” could now hold artificial intelligence back»⁴¹.

Che i propositi imitativi, anche alquanto narcisistici, costituiscano un impedimento e limitino davvero i risultati, pratici e teorici, ottenibili, soffocando la creatività necessaria al progresso?

This article contributes to heated debate around the multidisciplinary field of Artificial Intelligence by adopting a specific perspective: the translational point of view. Machine Translation was the first application of computer technology in the humanities, both because of the general interest in the study of human language, given its strong identity connotation, and because of its high formalization. Through an overview of the architectures that have followed, the paper focuses on the fruitful dialogue between practice and theory, which triggered modern reflections on the mechanisms of translation as a decision-making and negotiation process. These include: the limits of the interlingual transfer, the entropy of information, the anisomorphism of natural languages, the mutual influence of language and cognition, the relevance of extratextual information and speakers' cognitive horizons, the concept of textual fidelity. The deepening of these aspects opens up wider reflections, such as the thorny question of the relationship between natural and artificial intelligence, the essence of understanding, and the possible heuristic impediment constituted by imitation.

⁴¹ Johannes Bruder, *Infrastructural intelligence: Contemporary entanglements between neuroscience and AI*, «Progress in Brain Research», 233 (31 luglio 2017), cap. 5, p. 101-128, <<https://doi.org/10.1016/bs.pbr.2017.06.004>>.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Tra Bellezza e Pericolo: le trappole delle intelligenze artificiali nell'arte

Gianluca Garofalo

C.atWork Creative

Alla ricerca di una perfezione estetica a basso costo, il pericolo dell'omologazione e del condizionamento è concreto. Le intelligenze artificiali generano capolavori digitali o insidie visive? Quali sono le trappole che si celano dietro la creatività algoritmica?

Il professore Paolo Benanti¹, in uno dei suoi numerosi interventi sul tema delle intelligenze artificiali (AI), osserva che la lente convessa è stata la causa di una enorme rivoluzione, che ha visto nascere il telescopio e il microscopio, oltre che aver spostato l'Uomo dal centro dell'universo ad una posizione periferica e minimale. Analogamente, l'avvento delle intelligenze artificiali ci obbliga ad osservare la realtà attraverso un *macroscopio*. Non è più l'immensamente grande o l'immensamente piccolo a spostare l'Uomo dalla sua centralità, ma l'immensamente complesso.

Osserviamo ed utilizziamo una tecnica che opera attraverso calcoli complessi su una mole incommensurabile di dati. Il problema, tuttavia, è che non siamo affatto preparati a farlo. Siamo, quindi, davanti ad un nuovo scenario di tipo antropologico, oltre che sociale.

In questo articolo cercheremo di fare chiarezza, procederemo per punti cardine, costruiremo dei piccoli fari che ci aiuteranno a comprendere meglio cosa abbiamo di fronte. Cercheremo, prima di tutto, di capire cosa siano questi sistemi di calcolo che chiamiamo intelligenza artificiale.

Le intelligenze artificiali non sono intelligenti

I sistemi di calcolo in esame sono estremamente performanti nel computare un'impressionante mole di dati, molto più di quanto potrebbe qualunque essere

¹ Presidente della Commissione AI per l'informazione, docente presso la Pontificia Università Gregoriana, l'Istituto Teologico di Assisi e il Pontificio Collegio Leoniano ad Anagni. Oltre ai corsi istituzionali di morale sessuale e bioetica, si occupa di neuroetica, etica delle tecnologie, intelligenza artificiale e postumano. Ha fatto parte della task force Intelligenza Artificiale che coadiuva l'Agenzia per l'Italia digitale. È membro corrispondente della Pontificia Accademia per la Vita con particolare mandato per il mondo delle intelligenze artificiali. A fine 2018 è stato selezionato dal Ministero dello Sviluppo Economico tra i trenta esperti che hanno il compito di elaborare le strategie nazionali sull'intelligenza artificiale e in materia di tecnologie basate su registri condivisi e blockchain.

umano, ma non ne hanno nessuna coscienza. Non sono provvisti di discernimento. Non hanno la capacità di comprendere il senso dei calcoli che eseguono. La tesi del filosofo Luciano Floridi² è che si sia operato un divorzio tra l'abilità di agire e l'intelligenza necessaria per agire. In altre parole, pur senza cognizione di causa, attraverso un calcolo statistico tra miliardi di immagini, testi o dati, il sistema riesce a proporre una risposta alla nostra interrogazione sufficientemente soddisfacente. Questo perché la risposta è in linea con la media che consideriamo accettabile. L'impressione è che risponda in modo senziente, ma è solo un'impressione, appunto.

La narrazione legata al fenomeno è fuorviante

Utilizzare termini come *imparare*, *apprendere*, *decidere*, *addestrare*, altera la percezione reale del fenomeno. In effetti, nel momento in cui si prova a chiamare il *machine learning*, ovvero l'addestramento dell'intelligenza artificiale, con il suo vero nome, cioè calibrazione statistica del sistema di calcolo, la poesia svanisce.

Il fatto è che questi sistemi non imparano affatto, ma subiscono una regolazione per aumentare la precisione del calcolo, oltre ad un incremento dei dati a disposizione. Questi dati, tra l'altro, sono spesso coperti da copyright, ma tale realtà sembra assumere l'aspetto di un piccolo particolare irrilevante, soprattutto agli occhi di chi sostiene una visione dello sviluppo tecnologico che può essere definita religiosa, anche e, forse, principalmente, nell'espressione artistica e creativa dell'Uomo. Adottando la definizione proposta da Erich Fromm, la religione è qui intesa in senso lato, come «un sistema di pensiero e azione condiviso da un gruppo, che offre all'individuo un mezzo di orientamento e un oggetto di devozione»³.

Etichettatura o sfruttamento?

La macchina, ad un certo punto, senza un input specifico fornito dal tecnico, inizia a computare autonomamente l'enorme massa di dati provenienti da tutto ciò che, in qualche modo, è online. Il meccanismo di raccolta ed etichettatura può essere automatico, attraverso dei bot che raccolgono dati direttamente dalla rete, oppure manuale. I dati arrivano alla macchina solo dopo che un gran numero di persone, lavoratori del terzo mondo, il più delle volte sottopagati, ha "etichettato" i dati come conformi o meno. Proprio la necessità di rendere i dati conformi ad un criterio predefinito, sottopone queste persone ad un forte stress psicologico, per la necessità di escludere immagini non socialmente opportune come stupri, violenze, anche su minori, pornografia di ogni genere e tanto, tanto altro. Da un certo punto di vista, quanto accade non è poi così diverso dal colonialismo dei secoli tra-

² Professore ordinario di Filosofia ed etica dell'informazione presso l'Oxford Internet Institute della Oxford University, per la quale dirige dal 2017 il Digital Ethics Lab; chairman del Data Ethics Group dell'Alan Turing Institute.

³ Erich Fromm, *Avere o essere?*, Mondadori: Milano, 1976, p. 208.

scorsi, ovvero uno sfruttamento di lavoratori provenienti da paesi che hanno una scarsa capacità di tutela del lavoro. Niente di nuovo, purtroppo.

Da questo punto di vista, le scintillanti AI appaiono piuttosto opache.

Proprio questa necessità di conformità definita dall'azienda che produce il sistema, dà origine ad un pregiudizio molto condizionante, definito "pregiudizio algoritmico". Si tratta di una distorsione di partenza, una discriminazione che nasce prima ancora che i dati vengano raccolti e processati. Si tratta di un bias di tipo statistico. Se un dato è zero, non vuol dire che l'oggetto a cui fa riferimento non esista. Un esempio ampiamente riportato da diversi studiosi è quello legato alla previsione, da parte di una AI predittiva, di quale sarà il prossimo presidente degli Stati Uniti d'America. Il risultato ha dato indicazioni abbastanza precise: tendenzialmente di pelle bianca, sopra la sessantina, sicuramente uomo. La previsione è stata fatta sui dati del passato, sulle caratteristiche di tutti i presidenti eletti alla Casa Bianca. Dati acquisibili, non ipotetici, e non ci sono dati relativi a donne che abbiano ricoperto questo ruolo. In effetti, non è mai successo che una donna salisse alla Casa Bianca. Un altro esempio che spiega molto bene il pregiudizio algoritmico, è il caso della ragazza afgana. Utilizzando "ragazza afgana" nel prompt per ottenere un'immagine da un'AI generativa, il risultato è stato una quantità pressoché infinita di variazioni sul tema della famosa fotografia di Steve McCurry, l'ipnotica ragazza afgana dagli occhi verdi pubblicata nel giugno del 1985 sulla copertina del National Geographic. Tutte le ragazze afgane, per l'AI, hanno solo e soltanto quelle specifiche caratteristiche. Il dato trascritto con quel titolo coincide solo con quell'immagine, perché è talmente alto il grado di incidenza nel calcolo statistico che non ve ne possono essere altri. Ma noi sappiamo bene che di ragazze afgane ne esistono anche di molto diverse da quella raffigurata in quel meraviglioso ritratto.

Non solo. Si tratta di un problema che assume contorni preoccupanti anche in virtù di un altro gap, ossia quello delle persone senza competenze tecniche a cui è affidato il compito di etichettare i dati.

Il calcolo procede a livello sub-simbolico

Il sistema, dal momento in cui inizia a procedere autonomamente, opera un calcolo seguendo algoritmi che non sono più riconducibili ad una calibrazione di partenza e, quindi, non sono più così correggibili.

In altre parole, non siamo più in grado di collegare la reazione alla causa. Non sappiamo più comprendere perché la macchina dia determinate risposte, accettabili o inaccettabili che siano. Non abbiamo più il controllo.

Possiamo, a questo punto, considerare la macchina come una figura giuridica, che agisce autonomamente e di propria volontà? Possiamo sollevare l'umano, l'autore del prompt, dalla responsabilità di esserne l'artefice? A chi assegnare diritti e doveri? Il prompt designer è davvero l'autore dell'immagine e del testo generati? Chi è l'artista? L'Uomo decide o sceglie tra le decisioni della macchina?

Proprio in questo tipo di dinamiche si inserisce un fenomeno piuttosto noto: il feedback loop. È un fenomeno che si innesca nel momento in cui l'umano interagisce con l'AI. L'umano inserisce un input nel sistema e il sistema risponde. Fin qui sembra procedere tutto come ci si aspetterebbe. In seconda battuta, però, per calibrare il risultato, l'umano inserisce un nuovo input basato sulla precedente risposta della macchina, e la macchina risponde nuovamente. Dopo un certo numero di cicli è estremamente difficile capire se l'intenzione umana sia libera dal condizionamento del sistema stesso. In altre parole, non è più così facile capire chi decide, chi tiene il timone, se l'umano o la macchina. In questo loop non ha alcuna importanza il fatto che la macchina non abbia una volontà sua, il condizionamento è indipendente da questo, ma coerente con il bagaglio di dati che il sistema utilizza. Torniamo, quindi, ad essere condizionati da un pregiudizio algoritmico.

Da qui nasce il dubbio se uno strumento tecnico come questo sia davvero neutro e neutrale. Del resto, sappiamo bene che nessuna tecnica può definirsi neutrale.

La non neutralità della tecnica

Ecco un altro aspetto fondamentale per comprendere a fondo il motivo che rende ambiguo l'uso di questi sistemi. Scrive Martin Heidegger nel suo *La questione della tecnica*:

«Non possiamo quindi esperire veramente il nostro rapporto con l'essenza della tecnica finché ci limitiamo a rappresentarci la tecnicità e a praticarla, a rassegnarci ad essa o a fuggirla. Restiamo sempre prigionieri della tecnica e incatenati ad essa, sia che la accettiamo con entusiasmo, sia che la neghiamo con veemenza»⁴.

Hannah Arendt, nel suo *La banalità del male*, evidenzia come un'azione legata ad una narrazione efficace, anch'essa una tecnica, possa rendere accettabili e, addirittura, comportamenti virtuosi, le più terribili nefandezze. Nel testo, Arendt riporta le parole del gerarca nazista Adolf Eichmann:

«Hitler, disse, “avrà anche sbagliato su tutta la linea; ma una cosa è certa: fu un uomo capace di farsi strada e salire dal grado di caporale dell'esercito tedesco al rango di Führer di una nazione di quasi ottanta milioni di persone... Il suo successo bastò da solo a dimostrarmi che dovevo sottostargli”. E in effetti la sua coscienza si tranquillizzò al vedere lo zelo con cui la “buona società” reagiva dappertutto allo stesso suo modo. Egli non ebbe bisogno di “chiudere gli orecchi”, come si esprime il verdetto, “per non ascoltare la voce della coscienza”: non perché non avesse una coscienza, ma perché la sua coscienza gli parlava con una “voce rispettabile”, la voce della rispettabile società che lo circondava»⁵.

L'ermeneutica del manufatto tecnologico e la sua vera essenza evidenziano un da-

⁴ Martin Heidegger, *La questione della tecnica*, Firenze: GoWare, 2017.

⁵ Hannah Arendt, *La banalità del male*, Milano: Feltrinelli, 2023 (prima edizione digitale).

to essenziale: la tecnica risponde ad una domanda dell'Uomo. Questa domanda è ciò che l'Uomo immagina di dover avere per risolvere e migliorare la sua esistenza. In una condizione di benessere e fertilità intellettuale, l'Uomo inventa l'agorà, la filosofia e l'organizzazione sociale democratica ed etica. In una condizione di paura, malessere e pericolo, l'Uomo inventa la falange oplitica, le armi e la guerra. A quale tipo di società dobbiamo ascrivere l'invenzione delle AI?

Una considerazione interessante va fatta a riguardo delle AI che suggeriscono consigli sulle piattaforme di commercio online: «forse ti interessa anche quest'altro articolo...», «altre persone hanno acquistato anche...». Non è raro pensare che, effettivamente, la proposta sia interessante, ma ne abbiamo davvero bisogno? Avremmo scelto proprio quell'articolo se non avessimo avuto quel "consiglio"? Sembra un servizio su misura, ma a ben vedere è un'opera di omologazione.

La rivoluzione big data

Dopo aver inquadrato cosa siano i sistemi definiti Artificial Intelligence, cerchiamo ora di entrare un po' di più nella questione sociale e antropologica.

Sappiamo che l'Uomo nasce tecnologico, come specie.

Con il termine tecnica intendiamo cose, sistemi, anche molto diversi tra loro, ma che hanno tutti la stessa caratteristica: rispondere ad una domanda, ad una necessità dell'Uomo.

Così, per una domanda di comunicazione molto pressante, nascono un codice, un simbolo, un'icona e un linguaggio. Seguono sistemi per esprimere il senso, per trasportare le idee, per lasciare istruzioni, per superare l'assenza e indicare dei modi. Segno, disegno, scrittura, fino al nostro contemporaneo che ci consegna un fatto sorprendente: solo tra il 2013 e il 2015, l'umanità ha prodotto più dati che in tutta la sua storia⁶ ed è stata in grado di trasmetterli velocemente come mai prima. Come se tutta la produzione, dall'origine dell'umanità fino al 2012, tra romanzi, saggi, poesia, canzoni, musica, scienza, pittura, scultura, architettura, fotografia, cinema, tecnologia, non bastasse ad eguagliare la quantità di opere prodotte in soli due anni. Come se in questi due anni, per fare un paragone di tipo temporale, avessimo visto nascere il Partenone qualche ora prima del lancio dell'Apollo 11.

Questa enorme quantità di dati aumenta quotidianamente. Si è stimato che nel 2020 abbiamo prodotto circa 44 zettabyte di dati e li abbiamo fatti circolare istantaneamente per il globo, ovvero 440 miliardi di gigabyte alla velocità della luce⁷.

Com'è possibile?

In realtà, comprendere come questo sia avvenuto è abbastanza semplice: date uno

⁶ Bernard Marr, *The Complete Beginner's Guide to Big Data Everyone Can Understand*, «Forbes», (14 marzo 2017), <<https://www.forbes.com/sites/bernardmarr/2017/03/14/the-complete-beginners-guide-to-big-data-in-2017/>>.

⁷ Federica Brancale, *Big Data: Cosa sono, la storia, le caratteristiche, le analisi ed esempi*, Marketing Freaks, 2024, <<https://www.themarketingfreaks.com/2019/11/big-data-cosa-sono-la-storia-le-caratteristiche-le-analisi-ed-esempi/>>.

smartphone a tutti e inventate dei luoghi virtuali dove riversare parole, foto, video, disegni, citazioni, news, fake news, stati d'animo, diari personali, giochini di riconoscimento facciale, test psicologici, comizi, opinioni...

In due anni il gioco è fatto!

Il primo smartphone realmente tale, cioè come lo immaginiamo oggi, fa il suo esordio tra il 2008 e il 2009. Facebook, ciò che possiamo definire il social network più famoso e usato al mondo (stacca ancora di poco YouTube), l'erede dei forum, nato nel 2004 come realtà legata alla vita universitaria di Harvard, poi Stanford, Columbia University e Yale, Boston University e Boston College, proprio nel 2009, uscito dai confini dei campus, registra il suo primo anno in attivo. Solo in Italia, nell'agosto del 2008, Facebook registra un incremento di iscrizioni del 961% rispetto al mese precedente.

La domanda dell'Uomo

Solo un paio di decenni fa, quest'enorme quantità di dati avrebbe chiesto macchine estremamente costose, di grandi dimensioni, le risorse computazionali non sarebbero state adeguate, e avrebbe avuto bisogno di tanto, tanto tempo per essere analizzata, utilizzata e trasformata. Oggi possiamo farlo con un portatile, un tablet o uno smartphone, grazie all'innesto degli algoritmi capaci di trattare così tante variabili in poco tempo.

Possiamo digitare un prompt sulla tastiera e ritrovarci una splendida immagine sul monitor. Con due click mostrare al mondo il piatto che stiamo degustando in un ristorante nella periferia di Milano. Facendolo, produciamo dati ad una velocità assurda.

Ma possiamo davvero immaginare che la qualità delle pubblicazioni di Dante Alighiero, sul suo profilo Facebook, sia equiparabile alla Divina Commedia? O che il selfie di Maia Lino abbia un valore culturale degno di nota quanto l'autoritratto con passaporto ebreo di Felix Nussbaum?

Nel 2005 David Foster Wallace, profeticamente, pronunciò, davanti alla classe di laureati del Kenyon College, un discorso che in seguito diventò molto famoso e che, con una storiella spiritosa, individuava molto bene una caratteristica essenziale della nostra specie.

«Ci sono questi due giovani pesci che nuotano e incontrano per caso un pesce più grande che nuota dall'altra parte, che fa loro un cenno e dice: "Buongiorno, ragazzi. Com'è l'acqua?". I due giovani pesci nuotano per un po' e alla fine uno di loro guarda l'altro e dice: "Che diavolo è l'acqua?"»⁸.

⁸ David Foster Wallace, *Questa è l'acqua*, Torino: Einaudi, 2009. Così il grande scrittore americano introduceva il suo discorso di congratulazioni ai laureati del prestigioso Kenyon College nel 2005; un intervento divenuto famoso con il titolo *"This is water"* in cui l'autore di *Infinite Jest* spiegava ai laureandi il vero valore dell'educazione superiore, quella capacità di analizzare l'essenza della realtà che ci circonda così da costruire delle possibilità di scelta.

L'Uomo è una specie estremamente adattabile e questo può essere un pregio, ma anche un difetto. Dopo un certo tempo si abitua e smette di avvertire come estraneo, estraneo, qualunque dispositivo tecnico stia adoperando. Occhiali, magliette, calzini, linguaggio, penne, pennello, computer, smartphone, algoritmo.

Di fatto, questo è un processo di incorporazione che sfoca la visione della realtà sempre di più, via via che la distanza tra la mano e il *manicum* tecnologico aumenta. Via via che la tecnologia diventa più complessa e meno immediatamente riconoscibile e indagabile.

Sarà capitato a tutti di eseguire alcuni gesti automatici che, però, ci lasciano confusi e che hanno origine nel mondo digitale. Cercare di zoomare l'immagine di un libro, scrollare una fotografia per vedere se si riesce a scorgere altri dettagli sopra o sotto la cornice che la definisce, pensare di poter tornare indietro quando si sbaglia qualcosa, si lascia cadere un bicchiere, premendo contemporaneamente Control+Z.

Quello che abbiamo creato è un ambiente vitale a logica digitale e guidato interamente a livello digitale. Un ambiente che ci condiziona anche dove il nostro agire dovrebbe essere analogico, per così dire. Il nostro pensare è fortemente condizionato da questa logica virtuale.

Tutta questa massa di dati, di immagini, di opinioni, di post, che riversiamo nel web ogni minuto, da circa dieci anni, è davvero, come sostiene qualcuno, uno specchio, un'immagine fedele della nostra umanità? Della nostra società? Questi atteggiamenti sono davvero liberi?

Quando diamo alle AI generative di immagini il ruolo di *veicoli di democrazia nell'arte*, siamo davvero coscienti dell'azione che stiamo supportando? Questa visione postumanista e transumanista che vede la salvezza dell'umano, la sua liberazione, attraverso l'azione della tecnica, verso uno stato ideale di nuovo Eden, di «paradiso disincarnato del cyberspazio»⁹, è davvero l'evoluzione più auspicabile, la più giusta?

Possiamo sospendere ogni giudizio etico e accettare l'innovazione come progresso, senza nessuna azione? Passivamente?

Sono convinto che senza una opportuna educazione e senza un limite normativo umano, rischiamo di non essere più in grado di utilizzare e vivere l'ambiente digitale con consapevolezza e intenzionalità.

Il governo della tecnica sull'Uomo e non viceversa, si attua attraverso questo meccanismo di incorporazione. Il Control+Z da utile strumento si trasforma in *forma mentis*. L'azione non è più quella di un uso consapevole, di un governo umano dell'utente, ma, al contrario, un intervento della macchina a livello percettivo, interpretativo, neuronale.

⁹ Hava Tirosh-Samuels, *Transhumanism as a secularist faith*, «Zygon: Journal of Religion and Science», 47 (2012), n. 4, p. 710-733.

Allo stesso modo, la scelta di immagini generate da prompt linguistici, apparentemente governata dall'operatore, in realtà propone scelte determinate, condizionate dal sistema selettivo stesso della macchina, dall'azione dell'algoritmo.

L'azione arbitraria del calcolo avviene indipendentemente dal bagaglio esperienziale umano dell'operatore. Le *reference*, lo spazio culturale, emotivo, diventano aziendali e non personali.

Di fatto, potremmo trovarci davanti ad una sostituzione condizionata e arbitraria della cultura visiva individuale.

Per questi motivi, non credo sia corretto demonizzare una tecnologia innovativa come i sistemi di AI, ma, allo stesso modo, non credo sia opportuno porsi con un approccio religioso, mitologico, al tema. È imprescindibile vedere la natura ambivalente dello sviluppo tecnico, che non è mai scontato nell'imprevedibilità delle metodologie ed è sempre figlio della realtà politica, sociale e culturale contingente.

Bypassare le azioni umane di progettazione e, conseguentemente, di comunicazione, o limitarle a una scelta condizionata che mima un'arbitrarietà, ma che esclude la radice stessa dell'azione percettiva, interpretativa e reattiva del cervello umano all'origine, trasforma nella sostanza e non solo nella forma un sistema di comunicazione e di educazione, culturale e quindi sostanziale, tipico e peculiare della nostra specie.

Di fatto, abdichiamo alla nostra intelligenza, non soltanto alla nostra abilità di fare. Qual è la domanda alla quale questa tecnica risponde?

In the pursuit of low-cost aesthetic perfection, the danger of standardization and conditioning is real. Do artificial intelligences generate digital masterpieces or visual pitfalls? What are the traps hidden behind algorithmic creativity?

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024



Segnalazioni

Il progetto ADELE e il tool per l'autovalutazione digitale

Flavia Massara

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU)

Lo scorso febbraio si è concluso, dopo due anni di attività, il progetto europeo ADELE (Advancing Digital Empowerment for Libraries in Europe - 2021-1-IE01-KA220-ADU-000029436)¹. Il progetto finanziato dal programma Erasmus+ KA2 ha visto la collaborazione di istituzioni che avevano precedentemente collaborato al progetto BIBLIO (Boosting digital skills and competences for Librarians in Europe - 612411-EPP-1-2019-1-IT-EPPKA2-SSA)² ad eccezione del coordinatore di progetto: Local Government Management Agency (LGMA). I restanti partner sono stati: l'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), l'agenzia formativa European Grants International Academy Srl (EGInA Srl), la rete delle biblioteche in Bulgaria, Global Libraries - Bulgaria Foundation (GLBF) e Public Libraries 2030 (PL2030), l'organizzazione no-profit che si occupa delle biblioteche europee.

Il progetto ha avuto come obiettivo principale la realizzazione di un tool per l'autovalutazione delle competenze e delle strategie digitali delle biblioteche al fine di progettare attività di miglioramento. Tutto il lavoro svolto sarebbe stato impossibile senza il sostegno e il costante impegno della rete di 100 biblioteche pubbliche europee costituita all'inizio del progetto. Oltre all'Irlanda, l'Italia e la Bulgaria, Paesi del partenariato, hanno fatto parte della

rete biblioteche provenienti da Austria, Croazia, Germania, Grecia, Finlandia, Lituania, Polonia, Serbia, Slovenia, Svezia, Repubblica Ceca e Romania. I bibliotecari sono stati sin da subito coinvolti nella progettazione del tool e diversi sono stati i momenti di confronto per la realizzazione di uno strumento che fosse realmente in linea con le loro attività, i loro bisogni e le aspettative.

Lo strumento è attualmente disponibile online sul sito del progetto in tutte le lingue dei Paesi partner (italiano, inglese, bulgaro, gaelico e tedesco) insieme ad un prezioso kit di materiali creati per agevolarne l'utilizzo: il manuale utente, quattro webinar di formazione sul tema del digitale, una raccolta di 100 buone pratiche ispirata alle aree in cui è articolato il tool e altri materiali formativi.

Ciascuna biblioteca può utilizzare gratuitamente l'ADELE tool costruendo il proprio questionario di autovalutazione che può essere somministrato al massimo in due sessioni annuali per il monitoraggio. Il questionario è suddiviso in sei aree tematiche (Management; Infrastrutture, attrezzature, supporto e risorse; Sviluppo professionale continuo; Autovalutazione delle competenze digitali; Opportunità formative su competenze digitali per gli utenti; Collaborazione, networking e comunità) e prevede degli *statement* ai quali è possibile rispondere tramite la scala Likert; vi sono *statement* obbligatori, alcuni facoltati-

¹ Per maggiori informazioni consultare il sito web del progetto: <<https://www.adele-project.eu>>.

² Il progetto BIBLIO si è focalizzato principalmente sullo studio delle competenze necessarie per i bibliotecari contemporanei e ha prodotto due curricula di Istruzione e Formazione Professionale, IFP: "Community Engagement and Communication Officer" e "Digital Transformation Facilitator". Per la descrizione completa dei profili professionali vedere sul sito web del progetto: <<https://www.biblio-project.eu/>>.

tivi e la possibilità di creare nuove aree e nuovi *statement* per rispondere alle singole esigenze del proprio contesto bibliotecario. L'ADELE tool si rivolge a diversi profili contemporaneamente: il direttore della biblioteca, i bibliotecari e infine gli utenti, con l'obiettivo di riconsegnare un'immagine a tutto tondo dell'istituzione. La compilazione del questionario è anonima ma vi è la possibilità di rispondere a qualche domanda conclusiva sugli anni di esperienza nel settore e sul livello di competenze digitali.

A partire dall'estate 2023, il progetto ha previsto, inoltre, delle settimane di formazione internazionali (*Learning Teaching Training Week - LTTA*) nei quattro Paesi partner per le biblioteche coinvolte nel progetto. Questo momento d'incontro ha avuto un forte impatto sui professionisti dei diversi Paesi.

L'ICCU ha presentato i risultati del progetto ADELE al seminario "Sentire digitale: quali

competenze per i bibliotecari?" organizzato dal Laboratorio di Biblioteconomia sociale e ricerca applicata alle biblioteche (BIBLAB) presso Sapienza Università di Roma il 20 febbraio. L'evento è stato realizzato in collaborazione con l'Istituto, con il patrocinio dell'Associazione italiana biblioteche (AIB) - Sezione Lazio e con EGInA, partner di progetto.

La promozione dei risultati si è conclusa al convegno delle Stelline, appuntamento annuale dei bibliotecari italiani, il 20 marzo 2024, con il workshop: "Tecnologie digitali in biblioteca: come gestirle, come valutarle". In questo contesto l'ADELE tool è stato presentato a un'ampia platea di professionisti mostrando come, grazie alla sua flessibilità e personalizzazione, possa essere uno strumento utile non soltanto alle biblioteche di pubblica lettura ma anche alle biblioteche specialistiche, accademiche e scolastiche.



Figura 1. La presentazione del progetto ADELE al Convegno delle Stelline a Milano, 20 marzo 2024

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Un ponte tra scienza e società: i nuovi prodotti editoriali digitali del Dipartimento scienze umane e sociali, patrimonio culturale del CNR

Giulia Antonini

Consiglio nazionale delle ricerche (CNR)

Nell'ambito del riordino del Consiglio nazionale delle ricerche (CNR) del 2003¹, una delle novità più rilevanti per l'ente fu l'individuazione al suo interno di un'unica macroarea di ricerca a carattere interdisciplinare, comprendente le scienze umanistiche accanto a quelle sociali, economiche e giuridiche, denominata Dipartimento identità culturale. Contestualmente venne creato il Dipartimento patrimonio culturale, area che includeva l'insieme delle attività di ricerca riguardanti la gestione, conservazione, valorizzazione, utilizzazione e fruizione del patrimonio culturale materiale e immateriale.

La specificità di entrambe le strutture, fin dalla loro nascita, risiedeva nel carattere trasversale e interdisciplinare della loro offerta scientifica. Nove anni più tardi, nel 2012, presero avvio le procedure di fusione dei due dipartimenti nell'attuale Dipartimento scienze umane e sociali, patrimonio culturale (DSU)², le cui ricerche abbracciano il complesso delle scienze umane e sociali assieme al patrimonio culturale materiale e immateriale. Le cinque macroaree disciplinari che lo caratterizzano – patrimonio culturale, scienze storico-filosofiche, scienze sociali, ricerca cognitiva e linguistica, scienze giuridiche – comprendono discipline e saperi molto diversi tra di loro, ma uniti dalla comu-

ne esigenza di contribuire alla conoscenza, conservazione, valorizzazione e fruizione dell'identità e del patrimonio culturale.

La missione del Dipartimento consiste nella promozione della ricerca fondamentale e applicata, attraverso un approccio interdisciplinare finalizzato all'avanzamento della conoscenza, all'innovazione tecnologica e sociale e allo sviluppo di indirizzi di *policy*.

Gli ambiti scientifici del DSU sono sviluppati da quindici Istituti dislocati su tutto il territorio nazionale, le cui attività di ricerca, trasferimento tecnologico e divulgazione sono perseguite in stretta collaborazione con il sistema produttivo, le componenti sociali e le istituzioni di governo, a tutti i livelli.

Nella tabella alla pagina seguente viene illustrata l'organizzazione del DSU attraverso l'indicazione dei quindici Istituti afferenti suddivisi per macroaree disciplinari.

Le scienze umane e sociali e il patrimonio culturale hanno oggi un ruolo cruciale nel fronteggiare le emergenze e i cambiamenti che la società si trova ad affrontare; ad esempio, la diffusione dell'intelligenza artificiale nella vita degli individui comporta nuovi problemi di comprensione: le scienze umane e sociali in questo contesto sono determinanti non solo

¹ D.lgs. 127/2003, <<https://www.gazzettaufficiale.it/eli/gu/2003/06/06/129/sg/pdf>>.

² <https://www.cnr.it/it/trasparenza/delibere-cda/documento/89309/002020020.pdf>.

MACROAREE DISCIPLINARI	ISTITUTI DSU
Patrimonio culturale	Istituto di scienze del patrimonio culturale (ISPC)
Scienze storico-filosofiche	Istituto per il lessico intellettuale europeo e storia delle idee (ILIESI) Istituto per la storia del pensiero filosofico e scientifico moderno (ISPF) Istituto di storia dell'Europa mediterranea (ISEM)
Scienze sociali	Istituto di ricerca su innovazione e servizi per lo sviluppo (IRISS) Istituto di ricerca sulla crescita economica sostenibile (IRCRES) Istituto di studi sul Mediterraneo (ISMed) Istituto di ricerche sulla popolazione e le politiche sociali (IRPPS)
Ricerca cognitiva e linguistica	Istituto per le tecnologie didattiche (ITD) Istituto di linguistica computazionale "Antonio Zampolli" (ILC) Istituto di scienze e tecnologie della cognizione (ISTC) Istituto opera del vocabolario italiano (OVI)
Scienze giuridiche	Istituto di studi giuridici internazionali (ISGI) Istituto di informatica giuridica e sistemi giudiziari (IGSG) Istituto di studi sui sistemi regionali federali e sulle autonomie "Massimo Severo Giannini" (ISSIRFA)

nel fornire interpretazioni ai fenomeni collegati alla pervasività dei sistemi artificiali ma anche nell'implementazione efficace delle nuove tecnologie intelligenti. Ancora, lo sviluppo e l'attuazione di politiche sociali efficaci per costruire una società inclusiva e partecipata richiedono progressi nella comprensione degli individui e dei gruppi chiamati ad affrontare tali sfide. Analogamente, nell'ambito della transizione ecologica è necessario investire in processi che diano luogo anche a una trasformazione nel sistema dei valori, in nuovi sistemi di *governance* e, più in generale, in cambiamenti che riguardano i comportamenti individuali e collettivi.

Gli esempi sopra richiamati mostrano come il fine principale delle attività del DSU sia quello di creare una rete efficiente di trasferimento

di informazioni e competenze che possa fornire indicazioni di *policy* per rispondere alle sfide della società, contribuire a uno sviluppo economico-sociale sostenibile e più egualitario e misurare l'impatto degli sviluppi tecnologici e applicativi delle scienze dure sulla vita degli individui.

In quest'ottica, per aumentare l'impatto della ricerca scientifica nella società civile, risulta quanto mai necessario investire nella comunicazione e nella valorizzazione dei risultati e dei prodotti della ricerca.

In tale scenario si inseriscono le due nuove collane digitali del DSU, *Policy Brief* e *Working Paper*. Fondate a febbraio 2024 dal direttore del DSU Salvatore Capasso, le due collane saranno strumenti digitali utili alla comunità scientifica e alla società civile per la

condivisione della conoscenza e la valorizzazione del ruolo delle scienze umane e sociali e del patrimonio culturale.

I *Policy Brief* nascono con l'intento di ridurre il divario tra il mondo della ricerca e l'elaborazione delle politiche, offrendosi come strumento di informazione e indirizzo dei processi decisionali; più nello specifico, la collana vuole estrapolare dai risultati della ricerca informazioni e indicazioni scientificamente puntuali, utili a migliorare la comprensione di tematiche complesse, informare su nuovi approcci, suggerire e proporre soluzioni innovative su questioni di interesse comune nonché proporre raccomandazioni basate su dati oggettivi per affrontare sfide specifiche o sfruttare opportunità. Si tratta di pubblicazioni di alto profilo, i cui destinatari principali sono le istituzioni, i decisori politici e altri soggetti coinvolti a vario titolo nei processi decisionali su questioni rilevanti nelle politiche pubbliche e nel dibattito sociale.

Un *policy brief* si distingue per alcune caratteristiche fondamentali, tra cui la sintesi, che lo rende conciso e facilmente consultabile; la struttura che prevede tipicamente sezioni come l'*executive summary*, la chiarezza del linguaggio; la presentazione del problema; le opzioni politiche possibili; l'analisi di ciascuna opzione e le raccomandazioni o suggerimenti. Questi contributi sono sempre basati su evidenze, supportati da ricerche e dati solidi, citano fonti e studi rilevanti per sostenere le argomentazioni presentate. I *policy brief*, inoltre, sono utili strumenti di divulgazione poiché possono contribuire a veicolare informazioni rilevanti a un pubblico più ampio, rendendo questioni politiche e sociali complesse più trasparenti e accessibili agli organi di stampa e alla collettività.

I primi due numeri della collana sono prossimi alla pubblicazione: il primo, scritto da un

gruppo di ricerca dell'Istituto di ricerca sulla crescita economica sostenibile (IRCRES)³ del DSU, analizza il tema dei finanziamenti competitivi per ricerca e sviluppo in Italia. Il lavoro dedica la prima parte ad analizzare l'andamento degli strumenti finanziari competitivi nel decennio 2012-2021, soffermandosi su alcune criticità che ne hanno condizionato l'efficacia, come la modestia della dotazione finanziaria, la discontinuità della pubblicazione dei bandi e la limitata focalizzazione dei programmi di ricerca su obiettivi legati alle grandi sfide sociali e tecnologiche. Dopo la pandemia Covid-19 e la successiva istituzione del Piano nazionale di ripresa e resilienza (PNRR), l'immissione di stanziamenti finanziari senza precedenti nel mondo della ricerca ha introdotto cambiamenti significativi e lanciato sfide ambiziose. Il contributo, analizzando nel dettaglio gli strumenti finanziari del PNRR, offre delle raccomandazioni chiave per consolidare il nuovo percorso avviato dal Piano, non solo dal punto di vista delle risorse investite ma anche per alimentare in futuro un circolo virtuoso in cui la scienza generi nuove opportunità tecnologiche e innovazioni, che a loro volta creino le risorse per finanziare lo sviluppo di nuove conoscenze.

Il secondo numero è frutto di una ricerca portata avanti dall'Istituto di ricerche sulla popolazione e le politiche sociali (IRPPS)⁴ del DSU, in collaborazione con l'Università di Bergamo e l'Università di Roma Tre. Il contributo, dedicato ai brevetti, delinea la posizione dell'Italia nel panorama tecnologico mondiale, analizzandone i settori emergenti sulla base delle statistiche brevettuali. I dati analizzati, se da un lato sono incoraggianti, dall'altro confermano il ruolo secondario dell'Italia nello scenario europeo, con numero di invenzioni di gran lunga inferiore a quelle registrate in Francia o in Germania. Il lavoro offre delle raccomandazioni chiave per sostenere e svi-

³ www.irces.cnr.it.

⁴ www.irpps.cnr.it.

luppare le competenze tecnologiche italiane, sia nelle imprese che nel settore pubblico, puntando sull'integrazione tra le varie componenti del sistema di innovazione e lo sviluppo di strategie di trasferimento tecnologico tra università, enti pubblici di ricerca e imprese.

La collana *Working Paper* nasce con l'intento di offrirsi come strumento di discussione e approfondimento all'interno della comunità scientifica del DSU, nonché di diffondere i risultati di progetti e ricerche in corso.

La collana è prioritariamente orientata a ospitare lavori originali del personale di ricerca degli Istituti DSU, compresi assegnisti, dottorandi, borsisti e personale non strutturato coinvolto in progetti CNR. I contributi riguardano risultati di ricerca che si trovino in uno stadio di sviluppo idoneo alla presentazione a conferenze di rilievo nonché alla sottomissione a riviste scientifiche o volumi nazionali e internazionali.

Le due collane sono dirette dal direttore del Dipartimento e sono dotate di un Comitato scientifico e uno tecnico-editoriale. Per ogni proposta ricevuta, il Comitato scientifico veri-

fica preliminarmente la compatibilità del lavoro con gli ambiti disciplinari e le linee di ricerca del Dipartimento e, valutata la qualità scientifica dello stesso, decide se accettare o respingere la pubblicazione. Il Comitato tecnico-editoriale supporta il direttore e il Comitato scientifico, coordina e gestisce le attività editoriali, stimola la presentazione di proposte di pubblicazione, contribuisce a promuovere la collana all'esterno e concorre a fornire indicazioni per migliorarla.

Entrambe le collane avranno formato elettronico open access e saranno accessibili dal sito del DSU⁵ in formato pdf. È stata scelta la diffusione in open access per contribuire al cambiamento culturale in atto a livello europeo e internazionale verso la scienza aperta, in base al quale dati, prodotti e processi derivanti dalle attività scientifiche sono resi accessibili a tutti, a tutti i livelli della società.

Il CNR contribuisce al processo di transizione digitale con investimenti mirati a favorire lo sviluppo di sistemi per la gestione e conservazione di documenti e risorse digitali e la creazione di piattaforme open access a servizio della comunità scientifica.

⁵ www.dsu.cnr.it.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024

Il Geoportale Nazionale per l'Archeologia. Presentato il 26 gennaio 2024 un bilancio del progetto a sei mesi dalla pubblicazione

Valeria Boi

Istituto Centrale per l'Archeologia (ICA), Ministero della Cultura

Il 26 gennaio 2024 la Sala conferenze della Biblioteca delle Arti del Complesso monumentale del San Michele, a Roma, ha ospitato una giornata di studi dedicata al Geoportale Nazionale per l'Archeologia (GNA), a sei mesi dalla sua pubblicazione online. Il portale è stato realizzato dall'Istituto Centrale per l'Archeologia (ICA)¹ come punto di raccolta e condivisione dei dati esito di tutte le indagini archeologiche condotte in Italia, ed è liberamente accessibile in rete all'indirizzo <https://gna.cultura.gov.it/>.

Il GNA ha visto i suoi primi step progettuali nel 2018, con l'avvio del censimento dei dati raccolti nell'ambito delle procedure di archeologia preventiva e conservati presso gli archivi delle Soprintendenze Archeologia belle arti e paesaggio. Dopo una prima sperimentazione avviata presso gli uffici di tre sole regioni pilota, la raccolta dei dati ha coinvolto via via tutti gli uffici territoriali del Ministero della Cultura (MiC), mentre sul fronte dei contenuti si è evoluta verso una campagna di registrazione di tutti i rinvenimenti archeologici, sulla base dei dati ricavabili dalla documentazione

conservata presso gli archivi e talvolta già confluiti in carte archeologiche nei tradizionali formati cartacei, nell'intento di promuovere la conoscenza e l'accessibilità pubblica di un patrimonio documentale di difficile accessibilità e spesso totalmente inedito.

In pochi anni il GNA ha esteso la sua sfera di azione dai dati di archivio alla documentazione preliminare delle indagini di archeologia preventiva (d.p.c.m 14/02/2022), per divenire infine, nel marzo 2024, lo standard per il conferimento dei dati minimi di tutte le indagini sul campo svolte con la Direzione scientifica del MiC (cir. min, Direzione generale Archeologia, belle arti e paesaggio, n. 9 del 28/03/2024)². Questo approccio sempre più inclusivo è stato caldeggiato dalle Soprintendenze, interessate alla possibilità di disporre di un sistema agile, accessibile in rete, per la consultazione di un primo quadro di conoscenze sul territorio. Anche per le ditte e i professionisti del settore il Geoportale rappresenta uno strumento che semplifica le attività di raccolta dei dati e di prima lettura del paesaggio archeologico. L'ICA ha predisposto

¹ Responsabile Scientifico del Progetto è la Direttrice dell'ICA, Elena Calandra, coadiuvata da Valeria Acconcia, Valeria Boi e Annalisa Falcone; la progettazione del database e degli applicativi per l'inserimento dei dati è affidata a Francesco Marucci e Ada Gabucci, che svolgono inoltre un ruolo imprescindibile di supporto agli utenti. Le attività di pubblicazione e gestione del GNA sono condotte grazie alla collaborazione di tutto il personale dell'ICA e all'apporto di altri Istituti MiC, di enti pubblici e privati (<<https://gna.cultura.gov.it/chi-siamo.html>>); imprescindibile è inoltre il supporto di singoli colleghi, di ricercatori e professionisti (<<https://gna.cultura.gov.it/ringraziamenti.html>>).

² Per approfondimenti si veda <https://gna.cultura.gov.it/wiki/index.php?title=Istruzioni_operative>.

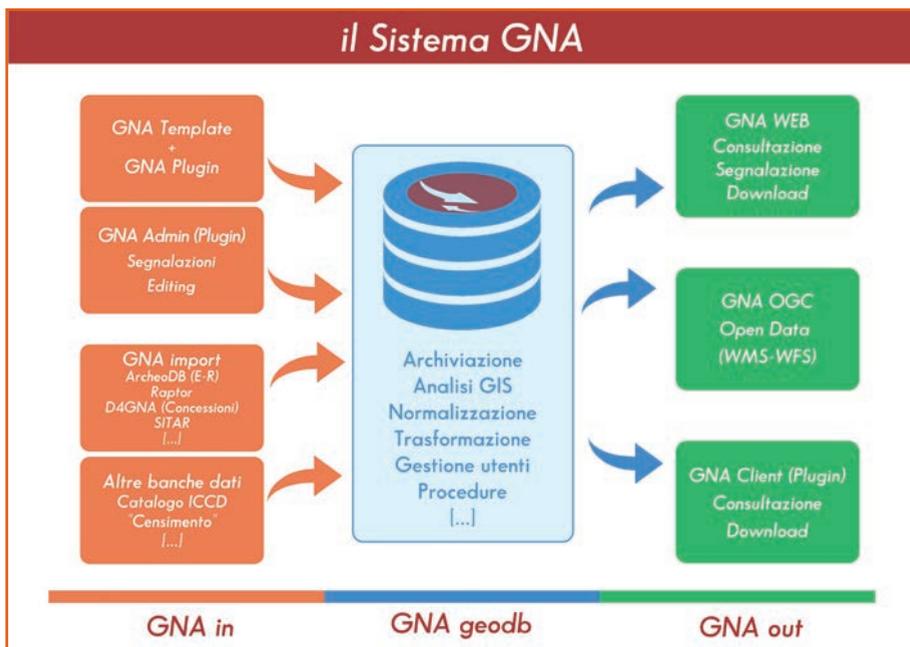


Figura 1. I componenti del Sistema GNA (Immagine di Francesco Marucci)

un template GIS per la raccolta dei dati, distribuito con licenza aperta, mediante il quale i dati già noti possono essere liberamente scaricati dal portale nella forma di un inventario redatto secondo lo standard GNA; l'operatore può modificare/aggiornare/correggere le schede già pubblicate e aggiungerne di nuove, inviandole poi al sistema mediante un plugin appositamente progettato (Fig. 1).

In fase di progettazione, si è ritenuto che questa soluzione fosse la chiave per garantire la sostenibilità del GNA: i dati sono registrati da chi li elabora e il loro caricamento sul GNA rappresenta uno degli adempimenti per la chiusura della documentazione scientifica da conferire al MiC. In questo modo, si limita la necessità di ricorrere a successive campagne di digitalizzazione della documentazione consegnata in formato cartaceo o comunque su supporti soggetti a rapida obsolescenza, quali CD, DVD, penne USB, il cui recupero compor-

terebbe tempi e costi più ingenti, oltre a possibili errori di trascrizione/interpretazione.

Le schede sono rilasciate con licenza CC BY e puntualmente attribuite al compilatore; se diverso da questo, come avviene nel caso di recupero di documenti d'archivio o bibliografia, è registrato anche il responsabile dei contenuti³. Si cerca, in questo modo, di dare massima visibilità e riconoscibilità alla componente autoriale, così da promuovere il ruolo del singolo operatore nella costruzione di questa base di conoscenza comune. Considerati i caratteri della banca dati e la sua modalità di implementazione, in cui ancora molte aree del territorio nazionale non sono rappresentate in maniera esaustiva, il GNA non sostituisce certo il dialogo puntuale con gli uffici competenti per la tutela del territorio; anche in futuro, esso non si sovrapporrà all'attività amministrativa degli uffici e al rilascio di pareri o nullaosta, come ribadito sul portale stesso.

³ <https://gna.cultura.gov.it/disclaimer.html>.

Sebbene con mero valore informativo, però, esso costituisce una utile base di partenza per lo studio: e l'“onere” di integrare il GNA con i dati ancora non presenti, condividendo il proprio lavoro, rafforza la consapevolezza che la conoscenza del patrimonio archeologico è attività collettiva, che prescinde dal ruolo e dalla specifica attività e che richiede un costante lavoro di aggiornamento e confronto, sebbene sul “terreno” virtuale di una carta archeologica digitale.

Con lo stesso obiettivo, il GNA recepisce direttamente i dati elaborati dai progetti ministeriali preesistenti: il progetto RAPTOR⁴ e l'ArcheoDB dell'Emilia-Romagna⁵, e a breve anche il SiTAR della Soprintendenza Speciale di Roma⁶. Allo stesso tempo, banche dati eterogenee e costruite in singoli contesti di ricerca sono in via di inclusione, grazie ad accordi stipulati caso per caso, e vengono pubblicate nella sezione “Altre banche dati”, valorizzando la componente topografica e la conoscibilità di progetti specialistici nati con strutture e per finalità diverse.

Già da tempo, in collaborazione con l'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU) e con il Polo Universitario Città di Prato (PIN), si sono inoltre sviluppate le attività derivate da un accordo siglato nel 2017, che vedeva l'ICA coinvolto nel progetto europeo Ariadne Plus⁷, in qualità di Affiliated entity dell'ICCU. Nell'ambito di tale iniziativa, e grazie alla partecipazione dell'Istituto di Scienza e Tecnologie dell'Informazione “Alessandro Faedo” - ISTI CNR di Pisa, è stata condotta una sperimentazione sulla gestione dei dati esito delle indagini archeologiche

svolte in regime di concessione ministeriale da università e altri enti di ricerca, che ha portato all'elaborazione del D4GNA, già presentato su questa stessa rivista⁸.

Nel momento in cui si chiude la stesura di questo contributo (6 maggio 2024), il GNA pubblica circa 37.500 rinvenimenti archeologici rappresentati come punti, poco più di 2000 features lineari e più di 17.000 features poligonali; ad essi si aggiungono più di 500 indagini svolte in regime di concessione e i 26.600 punti pubblicati nella sezione “altre banche dati”, ad oggi tutti esito del Progetto Censimento per la cartografia archeologica d'Italia, una fondamentale campagna di raccolta di dati bibliografici svolta fra il 2002 e il 2008 con il coordinamento della Sapienza Università di Roma e dell'Università di Foggia⁹. I numeri del Geoportale sono cresciuti esponenzialmente nei primi mesi successivi alla pubblicazione online, segno della crescente fiducia di ditte, professionisti e ricercatori e dell'aumento della sensibilità dei dipendenti del Ministero rispetto all'importanza del conferimento dei dati: in particolare, nei mesi successivi alla pubblicazione del GNA, numerose Soprintendenze hanno iniziato a richiedere spontaneamente a ditte e professionisti la consegna secondo lo standard GNA, anticipando di qualche mese l'inaugurazione di questa procedura a livello nazionale.

L'appuntamento del 26 gennaio scorso ha rappresentato una preziosa occasione di confronto e di scambio, che ha lasciato spazio, nel corso della mattinata, alla presentazione di alcune prospettive future per poi lasciare la parola, nel pomeriggio, a progetti che risultano già contributori attivi del GNA;

⁴ <https://raptor.cultura.gov.it/>.

⁵ <https://www.patrimonioculturale-er.it/webgis/>.

⁶ <https://www.archeositarproject.it/>.

⁷ <https://ariadne-infrastructure.eu/>.

⁸ Valeria Acconcia — Valeria Boi — Leonardo Candela — Annalisa Falcone — Francesco Mangiacrapa — Flavia Massara — Pasquale Pagano — Fabio Sinibaldi, *ARIADNE Plus e il D4GNA – Dataset per il Geoportale Nazionale per l'Archeologia*, «DigItalia. Rivista del digitale nei beni culturali», 18 (2023), n. 1, p. 129-140, <<https://doi.org/10.36181/digitalia-00064>>.

⁹ <https://gna.cultura.gov.it/collaborazioni.html>.

l'evento, che ha visto una grande partecipazione da parte del pubblico in presenza e attraverso la diretta trasmessa online, ha dunque rappresentato l'occasione per dare voce non solo a chi è stato coinvolto direttamente nelle fasi di progettazione ma, soprattutto, agli altri attori e testimoni del percorso, in una giornata di studio e confronto incentrata sui primi risultati raggiunti e sulle prospettive future di sviluppo.

L'introduzione ai lavori è stata affidata al Direttore Generale Archeologia, Belle Arti e Paesaggio Luigi La Rocca e ai Direttori degli Istituti centrali con i quali l'ICA si è rapportato, ossia l'ICCU, rappresentato dall'allora Direttrice Simonetta Buttò, e l'Istituto centrale per il catalogo e la documentazione (ICCD), diretto da Carlo Birrozzi: come già rammentato in precedenza, infatti, all'ICCU il GNA deve l'accordo con PIN – Vast Lab e CNR-ISTI, mentre l'ICCD ha collaborato attivamente alla definizione degli standard per la raccolta dei dati.

Le strategie, gli strumenti operativi e le funzionalità del GNA sono stati esposti nella sessione curata dall'ICA e dall'équipe di progettazione del portale: Elena Calandra in qualità di Direttrice, Valeria Boi in qualità di coordinatrice del progetto, con gli sviluppatori Francesco Marucci e Ada Gabucci, Annalisa Falcone per le indagini di ricerca, insieme con Pasquale Pagano, Francesco Mangiacrapa e Gian Luca Vannini del CNR ISTI; per gli aspetti editoriali e di formazione, infine, Valeria Acconcia, Domenico Marino, Miriam Taviani e Chiara Travaglini.

La sessione riservata agli sviluppi futuri, infine, è stata aperta da Laura Moro, oggi dirigente del Servizio II della Direzione Generale Archeologia Belle Arti e Paesaggio, con una riflessione sulle prospettive di integrazione dei dati geografici dei beni culturali, non solo archeologici.

Sono seguite la presentazione del progetto

Lazio Antico. Atlante del Lazio meridionale, con Paolo Carafa (Sapienza Università di Roma), l'esperienza dell'Università di Salerno con il progetto Coin Finds Hub, con Renata Cantilena, Giacomo Pardini e Federico Carbone, che hanno aperto la riflessione sulla connessione fra banche dati geografiche e i database dedicati ai beni mobili, e infine Emanuele Papi, Direttore della Scuola archeologica italiana di Atene, che ha superato i confini nazionali con la presentazione dell'applicazione degli standard GNA agli scavi italiani a Gortyna.

Nel pomeriggio le attività sono proseguite con il racconto della costruzione degli standard di interoperabilità fra il GNA e le piattaforme già esistenti, progettate dal Ministero della Cultura o esterne a esso, che raccolgono e archiviano i dati degli scavi archeologici; la sessione ha ripercorso le tappe della sfida rappresentata dalla costruzione di un linguaggio comune, che permette di salvaguardare e valorizzare specificità e operatività di esperienze diverse: sono intervenuti Matteo Frassine (Soprintendenza Archeologia, Belle Arti e Paesaggio per l'area metropolitana di Venezia e le province di Belluno, Padova e Treviso) e Stefania De Francesco (Soprintendenza Archeologia Belle Arti e Paesaggio per le Province di Bergamo e Brescia) per il progetto RAPTOR, Mirella Serlorenzi (Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma) per il progetto SITAR, Ilaria di Cocco per il Webgis dei Beni culturali dell'Emilia-Romagna e ArcheodB, e infine Maria Luisa Marchi (Università di Foggia) per il progetto Censimento e Gabriele Gattiglia (Università di Pisa) per il progetto Mappa.

Durante il Convegno, ulteriori applicazioni concrete del Geoportale sono state presentate grazie a una serie di poster, oggi consultabili online sul sito del GNA¹⁰.

Con la recente presentazione, il GNA ha se-

¹⁰ <https://gna.cultura.gov.it/materiali.html>.

gnato dunque una tappa importante all'interno del suo ormai articolato percorso di progettazione, svolto anche grazie alla collaborazione delle Soprintendenze, al contributo di altri enti di ricerca e degli archeologi che hanno partecipato alla sperimentazione. Non si può non sottolineare, infine, l'importanza del ruolo di tutte le ditte del settore, dei singoli funzio-

nari del MiC, professionisti e studiosi che quotidianamente condividono e segnalano criticità del sistema o propongono ottimizzazioni sia sul piano tecnico che nel modello dati: è grazie al loro supporto, infatti, che il GNA si può evolvere e migliorare per rispondere in modo sempre più efficace alle istanze per le quali è stata avviata la sua progettazione.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2024







Pubblicazioni

Futuro delle memorie digitali e patrimonio culturale: atti del convegno internazionale, Firenze, 16-17 ottobre 2003, a cura di Vittoria Tola e Cecilia Castellani, Roma, ICCU (2004)
ISBN 978-88-7107-110-7

Memorie digitali: rischi ed emergenze, a cura di Alessandra Ruggiero e Vittoria Tola, Roma, ICCU (2005)
ISBN 978-88-7107-111-5

Linee guida per registrazioni d'autorità e di rinvio, Roma, ICCU (2005)
ISBN 978-88-7107-105-0

Linee guida per la digitalizzazione del materiale fotografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale fotografico, Roma, ICCU (2005)
ISBN 978-88-7107-112-3

EAD: Descrizione archivistica codificata: dizionario dei marcatori, a cura di Giovanni Michetti, Roma, ICCU (2005)
ISBN 978-88-7107-115-8

MAG: metadati amministrativi e gestionali: manuale utente, a cura di Elena Pierazzo, Roma, ICCU (2006)
ISBN 978-88-7107-117-4

Linee guida per la digitalizzazione del materiale cartografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale cartografico, Roma, ICCU (2006)
ISBN 978-88-7107-116-6

Inter omnes: contributo allo studio delle marche dei tipografi e degli editori italiani del XVI secolo, Roma, ICCU (2006)
ISBN 978-88-7107-114-X

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 5: D, Roma, ICCU (2006)
ISBN 978-88-7107-113-1

Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti, a cura del Gruppo di lavoro sulla digitalizzazione di bandi, manifesti e fogli volanti, Roma, ICCU (2006)
ISBN 978-88-7107-118-2

Miniature e decorazioni dei manoscritti, a cura di Emilia Ambra, Angela Dillon Bussi, Massimo Menna, Roma, ICCU (2006)
ISBN 978-88-7107-119-0

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 6: E-F, Roma, ICCU (2007)
ISBN 978-88-7107-125-1

Il libro italiano del XVI secolo: conferme e novità in EDIT16: atti della giornata di studio, a cura di Rosaria Maria Servello, Roma, ICCU (2007)
ISBN 978-88-7107-124-4

OAIS: Sistema informativo aperto per l'archiviazione, a cura di Giovanni Michetti, Roma, ICCU (2007)
ISBN 978-88-7107-126-8

ISBD(CR): International Standard Bibliographic Description for Serials and Other Continuing Resources: revisione dell'ISBD(S), Roma, ICCU (2007)
ISBN 978-88-7107-104-2

Regole italiane di catalogazione: REICAT, a cura della Commissione permanente per la revisione delle regole italiane di catalogazione, Roma, ICCU (2009)
ISBN 978-88-7107-127-5

Requisiti funzionali per i dati di autorità: un modello concettuale, a cura di Glenn E. Patton, Roma, ICCU (2010)
ISBN 978-88-7107-129-9

International Standard Bibliographic Description (ISBD), IFLA, Roma, ICCU (2010)
ISBN 978-88-7107-128-2

Guida alla catalogazione SBN. Musica: musica e libretti a stampa, registrazioni sonore, video e risorse elettroniche musicali, Roma, ICCU (2012)
ISBN 978-88-7107-130-5

ISBD: International Standard Bibliographic Description, edizione consolidata, Roma, ICCU (2012)
ISBN 978-88-7107-131-2

Titolo Uniforme Musicale: norme per la redazione, Roma, ICCU (2014)
ISBN 978-88-7107-133-6



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
www.iccu.sbn.it

Copyright © ICCU - Roma

2024 - Istituto Poligrafico e Zecca dello Stato S.p.A. - Roma
Finito di stampare nel mese di giugno 2024
presso lo stabilimento IPZS di Via Salaria, 691 Roma